



US005712142A

United States Patent [19]

Adney et al.

[11] Patent Number: 5,712,142 [45] Date of Patent: Jan. 27, 1998

[54]) FOR INCREASING OSTABILITY IN CELLULASE IES
[75]	Inventors:	William S. Adney, Golden; Steven R. Thomas, Denver; John O. Baker, Golden; Michael E. Himmel, Littleton; Yat-Chen Chou, Wheat Ridge, all of Colo.
[73]	Assignee:	Midwest Research Institute, Kansas City, Mo.
[21]	Appl. No.:	604,913
[22]	Filed:	Feb. 22, 1996
	Rela	ated U.S. Application Data
[63]	No. 5,536,6 125,115, Se continuation No. 5,275,9	n-in-part of Ser. No. 276,213, Jul. 15, 1994, Pat. 55, which is a continuation-in-part of Ser. No. p. 21, 1993, Pat. No. 5,366,884, which is a -in-part of Ser. No. 826,089, Jan. 27, 1992, Pat. 44, which is a continuation-in-part of Ser. No. p. 26, 1989, Pat. No. 5,110,735.
[51]	Int. Cl. ⁶	
[52]		
[58]	Field of Se	earch
[56]		References Cited
	U.S	S. PATENT DOCUMENTS

5,110,735 5/1992 Tucker et al. 435/209

5,275,944	1/1994	Himmel et al	435/209
5,366,884	11/1994	Adney et al	435/209
5,432,075	7/1995	Himmel et al	435/209
5.536.655	7/1996	Thomas et al	435/209

OTHER PUBLICATIONS

Mohagheghi et al., Int. J. System. Bacteriol., 36:435-443 (1986).

Lejeune et al., Biosynthesis and Biodegradation of Cellulose, C. Haigler and P.J. Weimer, Ed., Marcel-Dekker, NY 1991, pp. 623-672.

Primary Examiner—Rebecca E. Prouty
Assistant Examiner—Tekchand Saidha
Attorney, Agent, or Firm—Edna M. O'Connor; Ken
Richardson; Ruth Eure

[57] ABSTRACT

The gene encoding Acidothermus cellulolyticus E1 endo-glucanase is cloned and expressed in Pichia pastoris. A new modified E1 endoglucanase enzyme comprising the catalytic domain of the full size E1 enzyme demonstrates enhanced thermostability and is produced by two methods. The first method of producing the new modified E1 is proteolytic cleavage to remove the cellulose binding domain and linker peptide of the full size E1. The second method of producing the new modified E1 is genetic truncation of the gene encoding the full size E1 so that the catalytic domain is expressed in the expression product.

13 Claims, 8 Drawing Sheets

U.S. Patent

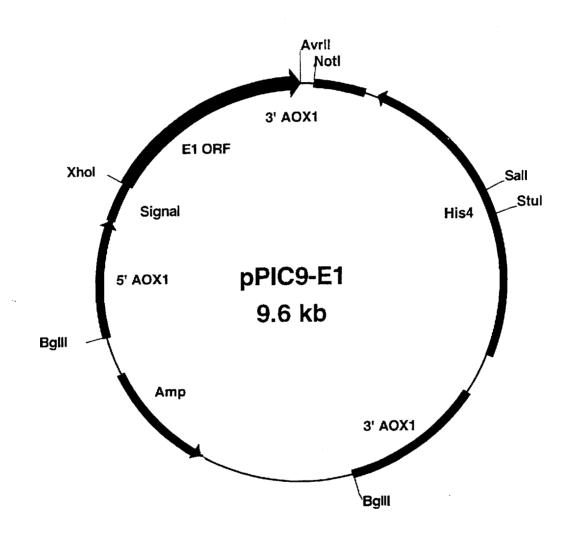
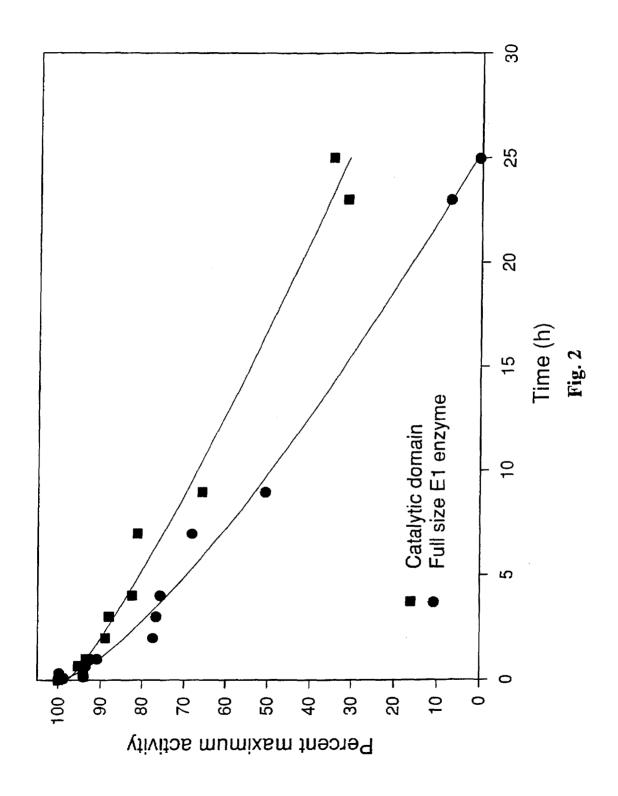
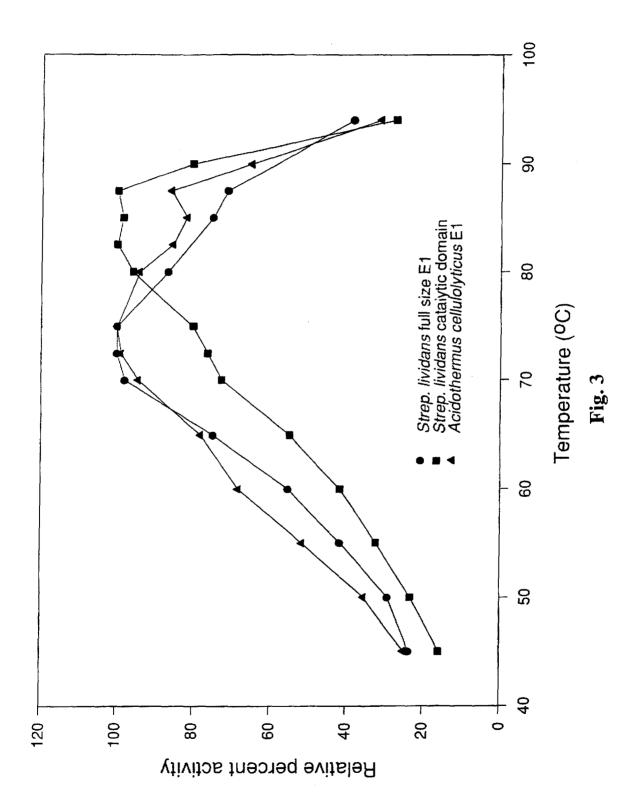
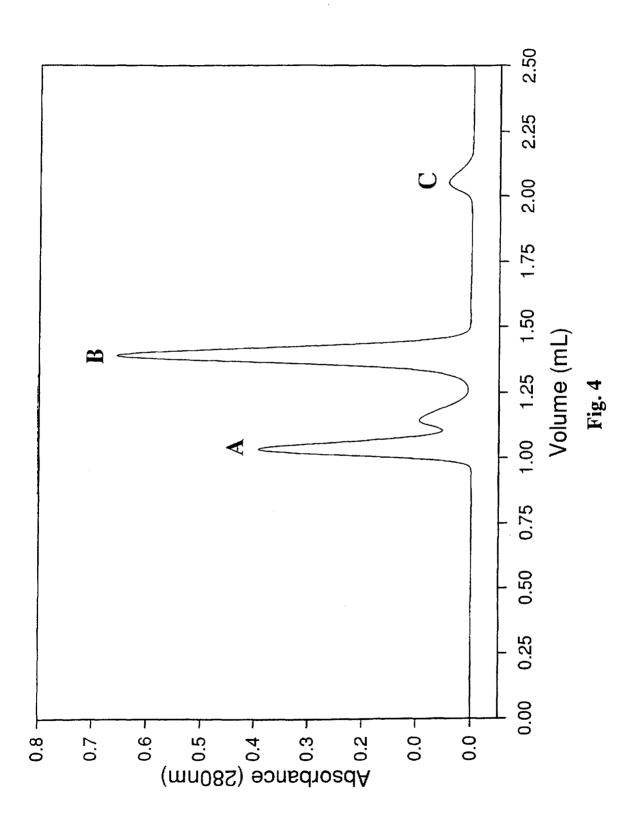


Fig. 1







Jan. 27, 1998

AGGGYWHTSG	REILDANNVP	VRIAGINWFG	FETCNYVVHG	LWSRDYRSML	50
DQIKSLGYNT	IRLPYSDDIL	KPGTMPNSIN	FYQMNQDLQG	LTSLQVMDKI	100
VAYAGQIGLR	IILDRHRPDC	SGQSALWYTS	SVSEATWISD	LQALAQRYKG	150
NPTVVGFDLH	NEPHDPACWG	CGDPSIDWRL	AAERAGNAVL	SVNPNLLIFV	200
EGVQSYNGDS	YWWGGNLQGA	GQYPVVLNVP	NRLVYSAHDY	ATSVYPQTWF	250
SDPTFPNNMP	GIWNKNWGYL	FNQNIAPVWL	GEFGTTLQST	TDQTWLKTLV	300
QYLRPTAQYG	ADSFQWTFWS	WNPDSGDTGG	ILKDDWQTVD	TVKDGYLAPI	350
KSSIFDPV					358

Fig. 5

GGATCCACCT TGTACAAGGT CACCTGTCC TGGTTCTGGT AGAGGGGGG GATGGTCACC GCGCACGATCT CTCCTTGTT GATGTCGACG GTCACGTGGT TACGGTTTGC CTCGGCCCGC ATTTTGCGCC TCGGGGTTGC TCCGGCTGTC GGGTTGGTT							
ATTITICGUE TEGGGETTEC TECGGETGET GGGTTEGGTT TGGCGTGGT TGGGGAGCAC GCCGAGGCGA TCCCAATGAG GGCAAGGGCA AGAGCGGAGC CGATGGCACG TCGGGTGGCC GATGGGGTAC CCCCAATGAG GCCAAGGGCCA AGAGCGGAC CGATGGCACG TCGGGTGGCC GATGGGGTAC GCCGATGAGG CGTGGCGTCC CCGCCCGGA CAGAACCGGA TGCGGAATAG GCCCTGTGAGC TGCCGGCTGG CGTTCGGATC ATGGGAACGA TCCCACCATT CCCCCGCAATC GCCCTGTGAGC TGCCGGCTGG CGTTCGGATC ATGGGAACGA TCCCACCATT CCCCCGCAATC GACCCGATCG GGACCAGGGC GGGCCGAGCC GGACCGTTGT GTCAATCCG CAAATCCAG CAATGCACC ATGGACAGGG ATTCTGACTC TGAGTAATCA TTGGATTCCC TCTTTCCCGC CTACCCGTTA ATGGACAGGG ATTCTGACTC TGAGTAATCA TTGGATTCCC TCTTTCCCGC CTACCCGTTA CCCAGAGTAG GCGACTGTAT GCCGTAAGTC GCGCATCCCG CACACCGTT TCCCCACAATC CCCAGAGTAG GCGACTGTAT GCGGTAGGTT GCCCCACACCCGAT TGGGATCGTT CCCCATAAGTT TCCGTCTCAC AACAGAACCA CGGGGGTTG GCGATCCCG GCGATTCCG GCTGGGAGA ACAGACGGG GAGAAACCAA CGGGGGATTG GCGGTCCGG GCGCATTCCG GTTGGCGCAA ACAGACGGG GAGAAACCAA CGGGGGGTTG GCGCTCCGG TGCTGCACTC GGTTGCCGCA TCCGCCACC TAGCCGTGCC GCCGGCCTC GTCTCCCGG TCCTGCCATT GGTTGCCCACAC TCGCCAACC TAGCCGTGCC GCGGCCCCCG GCGCATTCCG GTTGCCCACAC ACCCGCCCGC GCGCCCCCG GCGCCCCGG GCGCCGCTG CTTCGCCACAC ACCCGTCCC GCGCCCGCC GCGCCCCCG GCGCGCCTA CTCACACCG AGCGGCCCGA AAACCTGCAA TTACGTCGTG CGCGCCCCG GCGCGCCTA CTCACACCG ATCCTCACA CACCTGCAA TTACGTCGTG CACCGCTAC GCACCCCGC GCGCCCCGG CTCCCGCACC ATCCTCACA CACCTGCAA TTACGTCGTG CACCGCTAC CACCACCCGC CTCCCGCACC ATCCTCACA CTCCTCCCACC CACCACCCA CACCACCCC CACCACCCC CTCCGCCACC ATCCTCCACC CAGCCCCCC CACCACCCC CACCACCCC CACCACCCC CTCCGCCCC ATCCCCCACC CAGCCCCC CACCACCCC CACCACCCC CACCACCCC CACCAC	GGATCCACGT	TGTACAAGGT	CACCTGTCCG	TCGTTCTGGT	AGAGCGGCGG	GATGGTCACC	60
GCCGAGGGCA TCCCANTGAG GGCAAGGGCA AGAGCGGAGC CGATGGCACG TCGGGTGGCC 240 GATGGGGTAC GCCGATGGGG CGTGGCGCC CCGCCGCGA CAGAACCGGA TGCGGATAG 300 GTCACGGTGC GCCTGTGGAC CGTACCGCGA CAGAACCGGA TGCGGGGTC 360 GCCTGTGAGC GGCACGGCC GCGCGAACC CGCACATC CCCCCCAAT 420 GACGCGATCG GGCACGATC ATGGACAGGA CGCAGACCC GGCACGTGTG CCCAATACCCC 480 CATACGGTG GGCACTGTT TGTGAATCCC AGGCCAATACCC CCAATACCCC 540 ATGGACAGGA ATTGTGACTC TGAGTAAGCT TGCGACCACC AGCGTGACCCCC CAACACCGA 660 CTGGGACTC TTGACACCTC TAGCTGAACC GCGCCCCCC ATGATCAACC TGGGAGGCCGA GACATCCCCC 840 GCGAGTGCCC GCTCGCGGG TGATCACCGGG TGATCACCGGG TGATCACCGGG TGCTGCCCCC ATGATCACCCGG TGCTGCCCCC GCGCCCCCCC GCGCCCCCCC 840 GCGCCCCCCC GCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CGCACGATCI	CTCCTTTGTT	GATGTCGACG	GTCACGTGGT	TACGGTTTGC	CTCGGCCGCG	120
GATGGGGTAC GCCGATGGGG CGTGGCGTC CCGCCGCGGA CAGAACCGGA TGCGGGTC 360 GTCACGGTGC GACACTGTTGC CGTACCGGG ACCCGGATGA CAGAGGTGGG TGCGGGTGG 360 GCCTGTGAGC GACGCGAGC GGTCTGGATC ATGGGACGGG GCCCATCT 420 GACGCGATCG GGGCAGGC GGGCCAGCC GGGCCAGCT GTCAAGCCC ACGGTAGCC 480 ATGGACAGGG GCGCAGGC GGCCCATCT TGCAAGCCG CAAATGCCC CAAATGCAC 540 ATGGACAGGG GACGTGATT TGCGTAGCCC CCAAAACCGC CACACACCGA 660 CTGCGAGTT TCGGTAGCT GGCGTCCCCC ACAACACCC CGCGATGCCT TGGGATCCTT 720 CCCATAAGTT TCGCGACTCCAC ACACGACACC GGGGGGCCCC ATGGGGGGAGA ACAGAACCA GGGGGATCAC TGGGGACCCC AGCGGCCGGG TGAAACCACA CGGGGCCGGC GCGCCGCCCC ATGCCGGCCACA AGCGGCCACCA GCGCCGCCCC GCGCCCCCCC GCGCCCCCCC GCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ATTTTCGCGC	TCGGGCTTGC	TCCGGCTGTC	GGGTTCGGTT	TGGCGTGGTG	TGCGGAGCAC	180
GTCACGGTGC GACAGGTTGC CGTACCGCGG ACCCGGATGA CAAGGGTGGG TGCCGCTGG CGTCTGGATC ATGGAACGA TCCCACCATT CCCCGCAATC 420 GACGCGATCG GGAGCAGGCC GGCCCGAGCC GGACCGTGTG GTCGAGCCGG ACGATTCGCC 480 CATACGGTGC TGCAATGCC AGGCCCATGT TGTCAATCGC CCAAAGCCGG CAAGGCCACC 540 ATGGACAGG ATTGTGACTC TGGGATATGC CTGAGTATGC CTACGCGTTA 660 CCCAGAGTAG GCGGACGTT GGGCCCCCC CCAACACCGA TGGAGTCCTG 660 CCCATAAGTT TCCGTCAACC TGGTGACCCC CAACAACCGA TGGAGATCCTG 7720 CCCATAAGTT TCCGCCACC AACAGACGGG GAGAAACCAA CGGGGGGTCGCCGCG GCGCATTGCC 840 CCGATGCCCT GCGTGCCCCC ATGACTCCCG GCGCCCGCCC GCGCCCCCC GCGCACCAT 900 GGTTGCCCCC AGCGCCCGGG AGATCCTCGA GCGGACCACC GTCGCCCGCG GCGCCCCCCG GCGCCCCCGG 1020 CTACCCACCA ATCTCACACC AGCTCCCCACA GCCCACCACC<	GCCGAGGCGA	TCCCAATGAG	GGCAAGGGCA	AGAGCGGAGC	CGATGGCACG	TCGGGTGGCC	240
GCCTGTGAGC TGCCGGCTGG CGTCTGGATC ATGGAACCA TCCCGCCATC 420 GACGCGATCG GGAGCAGGGC GGCCGAGCC GGACCGTGTG GTCGAGCCGG ACGATTCGCC 480 CATACGGTGC TGCAATGCCC AGCGCCATGT TGTCAATCCG CCAAATGCAG 540 ATGGACAGG ATTGTACCTC TGAGTATCA TTGTCAATGCC CTATGCGCATC CGAAGATGC GCCAATGCTC 660 CGCAGAGTAG GGACTGTCA TGGTGAACTC CCAACACCGA TGGGATCGTT 720 CCCATAAGTT TCGCCTCAC AACAGAACGA CGGGGCCCCC ATGATCACG 784 ACGGGGGGA ACGACGGGG GAAAACCAA CGGGGCGCC CGCGTGCCC 684 GCGAGTGCCT GCCTCGCGG TGATGCTCCC CGCGCCGCG GCGCACTCC 840 GCGAGTGCCA ACGCCCCAC TAGCCGTCC GCGCCGCGC GCGCCGCGG TGCTGCCACC 780 GGTTGCCCCA ACGCCCCACC AGATCCCCGCA ACGCCCCCGC GCGCCCCCC GCGCCCCCC GCGCCCCCC GCGCCCCCC GCGCCCCCCC GCGCCCCCCC GCGCCCCCCC ACCC	GATGGGGTAC	GCCGATGGGG	CGTGGCGTCC	CCGCCGCGGA	CAGAACCGGA	TGCGGAATAG	300
ACCEGERATEG GRAGEAGGGC GGCGCGAGCC GGACCGTGTG GTCGAGCCGG ACGATTCGCC CATACGGTGC TGCAATGCCC AGCGCCATGT TGTCAATCCG CCAAATGCAG CAATGCACA ATGGACAGGG ATTGTGACTC TGAGTAATGA TTGGATTGCC TTCTTGCCGC CTACGCGTTA ATGGACAGGG ATTGTGACTC TGAGTAATGA TTGGATTGCC TTCTTGCCGC CTACGCGTTA CGCACAGAGTAG GCGACTGTAT GCGGTTGAACG CGCAATACTC CCAACACCGA TGGGATCGTT CCCATAAGTT TCGACACGTC TGGTTGAACG GCGCATACTC CCAACACCGA TGGGATCGTT ACCGGGAGCA ACAGACGGG GAGAAACCAA CGGGGGATTG GCGTCCAG TGGAATGGGT ACCGGGAGAA ACAGACGGG GAGAAACCAA CGGGGGATTG GCGTCCCGC GCGCATTGCG GCTGCCGAA CCAGCCGGG TGATCCTCCG GTCGGCCCCC ATGACCCGA TGGCATTCC GGTTGCCGCA CTCGCCAGC TAGCCGTGC GCGCATACCC GCGCCATTCCG GGTTGCCGCA CTCGCCAGC TAGCCGTGC GCGCACCAC GCGCATTCCG GGTTGCCGCA CTCGCCACC TAGCCGTGC GCGCGCGCT GTCGCCGGG TCCTGGCATT GCGACACCG AGCGCCCGG AGATCCTGCA CGCGCACCAC GCGCACCAC GCGCACCAC GCCCAACACC GCGCACCAC GCCCAACACC GCCCACCAC ATCCCCAC CCCCACACCAC ACCACACCC CACCAACACC CACCAACACC CACCAACACC CACCAACACC ACCACACCC CACCAACACC ATCCACTTT ACCAGATCGA 1140 TCAGGACCCTG CAGGGCTCA CTCTCCAC GGCCACCAC GCCCACCACC GCCCACCACC GCCCACCAC CCCCACCACC GCCCACCAC CCCCCCCC	GTCACGGTGC	GACATGTTGC	CGTACCGCGG	ACCCGGATGA	CAAGGGTGGG	TGCGCGGGTC	360
CATACGGTGC TGCAATGCCC AGCGCATGT TGTCAATCCG CCAAATGCAG CAATGCACC ATGGCAGGA ATGGACAGG ATTGTGACTC TGAGTAATGA TTGGATTGCC TTCTTGCCGC CTACGCGTTA 600 CGCAGAGGTAG GCGACTGTAT GCGGTAGGTT GGCGCTCAG CCGTGGGCTG GACATGCCTG 660 CTGCGAACTC TTGACACGTC TGGTTGAACG CGCAATACTC CCAACACCGA TGGGATCGTT 720 CCCATAACTT TCCGTCTCAC AACAGAATCG GTGCGCCCTC ATGACACCGA TGGGATCGTT 720 ACGGGGGAAACCAA CGGGGGGTTG GCGGCCCTC ATGACACCGA TGGGATCGTT 720 ACGGGGGAACACCA ACAGACACAC CGGGGGGTTG GCGCGCCGCG GCGCATTGCG 840 CGGTGCCGCG GCGCCCCCG GCGCACTTGCG GCGCACGCGG TGATGCCGG GAGAACCAA CGGGGGGTTG GCGCCCCCG GCGCCGCGG GCGGCGCTA 960 GCGGTGCCGC ACGCCCCACC TAGCCGTGCC GCGCCCGCG GCGGCGCAT 960 GCTGCCGCAAC ACGCGCGGG AAACCCAA TTACGTCCGC GCGCCCGCG GCGGCGGCTA 960 CTACCCAACC ACGCGCCGCG AAACCCAAC GCGCCGCGG GCGGCGCTA 960 CTACCCAACC ACGCCCCAA ACACCAACC GCGCCCGCG GCGGCGCTA 960 CTACCCGCAC ATGCCCGGA AAACCCTAA TTACGTCCGT CACGCGTA GGATCACCGGA 1080 CTACCCGAC ATGCCTGCA AAACCTACAC GCTCCACCA ACACCAACC GCGCCGCGA ACACCAACC GCCGCACCAC ACCACCACC ATCACTCCAA ATCCCCACA ACCACAACC ATCACTCTT ACCAGATCAA 1200 CTCCGGCCGC ATGCCCGCA AGACCACAC ATCACTTTT ACCAGATGAA 1200 CTCCGGCCGC CTGCGCACCA ATCACTTTT ACCAGATGAA 1200 CTCAGGACCG CTGCGCATCA TTCTTGACCC CCACCGACCG GATTGCACCG GCACCTCCGC 1320 CCTGCTGGTAC AAGGGACACC CTCCGGAGCC ACCTGACCG GATTGCACCG CGACCGCCG ATGCCCGCC AACGCCGCC ACGACCACCA ACACCAACC ACGCCGCCC ACGACCACCA ACGCCTACCACC ACCGCCCGCC ACGACCACCA ACCGCCCGCCC	GCCTGTGAGC	TGCCGGCTGG	CGTCTGGATC	ATGGGAACGA	TCCCACCATT	CCCCGCAATC	420
ATGGACAGG ATTGTGACTC TGAGTAATGA TTGGATTGCC TTCTTGCCGC CTACGCGTTA CGCAGAGTAG GCGACTGTAT GCGGTAGGTT GGCGCTCCAG CCGTGGGCTG GACATGCCTG CTGCGAACTC TTGACACGTC TGGTTGAACG CGCAATACTC CCAACACCGA TGGGATCGTT 720 CCCATAACTT TCCGTCTCAC AACAGAATCG GTGCGCCCTC ATGATACAC TGAAAGGACT 780 ACGGGGGAGA ACAGACGGGG GAGAAACCAA CGGGGGATTG GCGGCTGCCG GCGCATTGCG 840 CGGATGCCT GGCTCGCGGG TGATGCTGCG GTCGGCGCT CTCGGCGGCT TGCGGGCTGCCG TGCTGCCGAC TGGCGCACC TAGCCGTGCC GCGCGCGCGCGCGCG TGCTGCCGAC TTGGCCGAC TAGCCGTGCC GCGGCGCGCT CGCGCCGCGG GCGGCGCTA 960 CTTGGCCACA ACGCGCAGC TAGCCGTGCC GCGGCAACAAC GTGCCGGTAC GGATCGCCGG 1020 CATCAACTGG TTTGGGTTCG AAACCTACA TTACGTCGTG CGCGCCGCG GCGGCGCTA 960 CTACCGCAGC ATGCTCGAC AGATAAAGTC GCTCGGACAAC GTGCCGGTAC GGATCACCGG 1020 CTACCGCAGC ATGCTCGAC AGATAAAGTC GCTCGGCTA AACACAATCC GCTCGCGTA 1140 CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAAC ATCAATTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCC CGTACGCCGG 1260 CTCAGATCGGC CTCCGCATCA TTCTTGACCG CCACCACCG GATTGCACGC GCACCTCGC 1320 CCTGTGGTAC ACGGGACACC TCTCGGAGGC TACGTTGAC CAAAATCGTCC CGTACGCCGG 1260 CCCGGCCTGC CGGGCTACC TCTCGGAGGC TACGTGGAT TCCGACCTG AAAGCCTTGCA 1380 CCCGGCCTGC TGGGGCTGC GCGACCGAC CATCAACTTT TCCGACCTG AAGCGCTGGC 1320 CCGGAAACGC CTCCTCTCGG TGAATCCGAC CATCGACTG CGATTGCACC CCGACCGGCT 1380 CCCGGCATCC TGGGGCTGC GCGACCCGC CATCGACCG GATTGCACC CCGACGGGC 1500 CGGAAACGC CTCCTCTCGG TGAATCCGAA CCTGCCTATT TTCGTCAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGCTCCG GCGACCCGC CACCCGCCG CGATGGCC CCGACGGGCC 1500 CGGAAACGC CTCCTCTCA ACCAGACCT CCCCACCACC GACTCGCC CAGCGCGCC 1500 CGGCAACCG TGCTCTCAC ACCAGACCT CCCCACCACC GACCGCC AGTACCCGGC 1500 CGGCAACCA TCCCGACCC ACCAGCACG CTCCTCCAC ACCACCAC ATCCCGCC ACCACCACCAC ATCCCGCC ACCACCACCAC ATCCCGCC ACCACCACCAC ATCCCGCC ACCACCACCAC ATCCCGCC ACCACCACCAC ATCCCGCC ACCACCACCAC ACCACCACCAC ATCCCCACCAC ACCACCACCAC ACCACCACCAC ACCACC	GACGCGATCG	GGAGCAGGGC	GGCGCGAGCC	GGACCGTGTG	GTCGAGCCGG	ACGATTCGCC	480
CCCACAGATAG CCGACTETAT GCGGTAGGTT GGCGCTCCAG CCGTGGGCTG GACATGCCTG CTGCGGAACTC TTGACACGTC TGGTTGAACG CGCAATACTC CCAACACCGA TGGGATCGTT 720 CCCATAAGTT TCCGTCTCAC AACAGAATCG GTGCGCCCTC ATGATCAACG TGAAAGGAGT 780 ACGGGGGAA ACAGACGGG GAGAAACCAA CGGGGGATTG GCGGTGCGCG GCGCATTGCC 840 CGGAGTGCCT GGCTCGCGG GAGAAACCAA CGGGGGATTG GCGGTGCGCG GCGCATTGCC 640 CGCGAGTGCCC GCGCACACCC TAGCCGTGCC GCGCGCGCG GCGCGCGCTA 960 CGTGGCCACC ATGCCGACC TAGCCGTGCC GCGCGCGCGC GCGCGCGCGCACACC ATGCCCACC AGCCCTCGCA ACCCTCGCAACC TAGCCGTGCC GCGCGCGCT GCCGCGCGG GCGCGCGCTA 960 CTCCCCAACCG AGCCGCCGG AGATCCTCGA AACCTGCAA TTACGTCTGC CACGGTCTCT GGTCACCGG 1020 CTCCCCACCC AGATCAACTGC AACCTCCAA TTACGTCTGC CACGGTTCC GGTCACCGGA 1080 CTCCCGACCA ATCTCCAACC AGATAAAGTC GCTCGCCAACAAC GTCCCGCAC ATCACACTTC AACCAATCC GCTCGCCGTA 1140 CTCTGACCAC ATTCCAACC AGATAAAGTC GCCGAACAAC ATCAATTTTT ACCAGATGAA 1200 CTCTGACCACC ATCTCAACG CGGCCACCAC GCCGAACAAC AACAATCC GCTGCCCGTA 1140 CTCTGACCAC ATCTCAAGC CGGCCACCAC GCCGAACACC AACAATCC GCTCCCGCA 1260 CTCGAGCCGC CTCGCGCAC ATCAACTCT ACCAGATGAA 1200 CTCGAGTCGCC CTGGCCTAC ATCTTCAACCG CCGCCACACAC AACAATCC GCCGCAGCGGC 1260 CTCGAGTCGCC CTGGCCTAC ATCTTCAACCG CCGCCACACAC AACAATCC GCCAGCCGCG 1320 CCTGCTGGTAC ACGAGCACC CTCCGGAGCC CACCGACCG GATTGCAACC GCCAGCGCCG CTACGCCGC CTACGCCGC CTACGCCGC CTGCGCCTAC AACAATCCTC CACGACCCG CTACGCCGC CTACGCCGC CTACGCCGC CTACGCCGC CTACGCCGC CTGCGCCTAC AACACACACC ACGCCGCCC CACGCCGCC AACACCCGCC CCGACCGCC CACCACACC CCCCAACAC ATGCCCGCC ACGCCCCACACC CCCCAACAC ATCCCCGCA CCCCACACAC ATCCCCCAACAC ATCCCCGCA CCCCACACAC ATCCCCGCA CCCCACACAC ATCCCCGCA CCCCACACAC ATCCCCGCA CCCCACACAC ATCCCCCAACAC ATCCCCGCA CCCCACACAC ATCCCCGCA CCCCACACAC ATCCCCGCA CCCCCACACAC ATCCCCGCA CCCCCACACAC ATCCCCGCAC CCCCACACAC ATCCCCCAACAC CACCACACAC ATCCCCCAACAC CACCACACAC CACCACACAC CACCACACAC CACCAC	CATACGGTGC	TGCAATGCCC	AGCGCCATGT	TGTCAATCCG	CCAAATGCAG	CAATGCACAC	540
CTGCGAACTC TTGACACGTC TGGTTGAACG CGCAATACTC CCAACACCGA TGGGATCGTT CCCATAAGTT TCCGTCTCAC AACAGAATCG GTGCGCCCTC ATGATCAACG TGAAAGGAGT ACGGGGGAGA ACAGACGGGG GAGAAACCAA CGGGGGATTG GCGGTGCCGC GCGCATTGCG B40 CGCAGTGCCT GGCTCGCGGG TGATGCTGCG GGTCGGCGT GTCGTCGCGG TGCTGCATT GGTTGCCGCA CTCGCCAACC TAGCCGTGCC GCGGCGGCT CGCGCCGCGG GCGGCGGCTA TTGGCACACG AGCGGCGGG AGATCCTGGA TTAGCTGTC CACGGTTCT GGTCACGCG TCTGCCACC CTACCACTG TTTGGTTCC AAACCTGCAA TTACGTCTT CACGGTTCT GGTCACGCGA CTACCGCAGC ATGCTCGACC AGATAAAGTC GCTCGGCTAC AACACAATCC GTCTGCCGTA TTAGGACCTG CAGGTCTGA CGTCCTTGCA GCTCGGCTAC AACACAATCC GTCTGCCGTA TCAGGACCTG CAGGTCTGA CGTCCTTGCA GGTCAGCGC ATCATTTT ACCAGATGAA TCAGGACCTG CAGGTCTGA CGTCCTTGCA GGTCATGGAC AAATCGTCC CGTACGCCGG TCAGATCGGC CTGCGCACA TTCTTAGACC CCACCGACCG GATTCACACG GGCAGTCGGC TCAGATCGGC CTGCGCACA TTCTTAGACC CCACCGACCG GATTGCACGC GGACGCCGG TCAGATCGGC CTGCGCATCA TTCTTAGACCG CCACCGACCG GATTGCACGC GGACGCCGG TCAGATCGGC CTGCGCATCA TTCTTAGACCG CCACCGACCG GATTGCACCG GGCAGTCGGC TCAGATCGGC CTGCGCATCA TTCTTGACCC CCACCGACCG GATTGCACCG GGCAGTCGGC TCAGATCGGC CTGCGCATCA TTCTTGACCC CCACCGACCG GATTGCACCG GCAGCGCGC 1320 GCCGGCCTAC AAGGGAACC CGACGGTCGT CGGCTTTGAC TTGCACAACG AGCCGCTGC 1380 GCAGCGCTAC AAGGGAACC CGACGGTCG CACCGACGC GATTGCACCG AACCGCTGC 1380 CCCGGCCTGC TGGGGCTGC GCGATCCGAG CATCGACTG CAGCGCGC CGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAC GTGGCAGAG 1560 CTACAACGGA GACCTCTCCT GCTGGGGCG CAACCTGCAA GGACCCGGCC AGTACCCGGT 1620 CCGCGCACAC GTGCCCAACC GCCTGGTGTA TCGCACCACAC ATGCCCGGC ACTGCCACAC GACCGCCTCA ACCTGCAA CACCGCACAC GCCCGCACACAC GCCCGCACACAC GCCCGCACACAC GCCCGCACACAC ATGCCCGGA TCTGGAACAA 1740 GAACTGGGGA TACCTTCTA ATCAGAACAT TGCACCGGTA TGGCCGGCA TCTGGAACAA 1740 GAACTGGGG TACCTGCAG ACCGGACCT GCTGAACACA TTCGGACCAA TACCGGCC ACCAGACACA TTCGAACACA TTCCACACAC TCAAGACCT TCCACACAC TCGCCCGCA TCTGCACAC ATCCGGCC ACCACACAC TCCACACAC TCACACACA TACCACGCC ACCAGACACA TACCACGCC ACCAGACACA TACCACGCC CCCACACAC TCCACACAC TACCACACC TCCCCCACACC CCCCACACCC TACCTGCCC CCGACCCCA ACCCCGCC ACCGCCCA ACCCCGCCA CCCCCCACACCC C	ATGGACAGGG	ATTGTGACTC	TGAGTAATGA	TTGGATTGCC	TTCTTGCCGC	CTACGCGTTA	600
CCCATAAGTT TCCGTCTCAC AACAGAATCG GTGCGCCCTC ATGATCAACG TGAAAAGGAGT 780 ACGGGGGGAA ACAGACGGGG GAGAAACCAA CGGGGGATTG GCGGTGCCGC GCGCATTGCC 840 GCGAGTGCCT GGCTCGCGGG TGATGCTGCC GGTCGGCGT GTCGTCGCGG TGCTGCCATT 900 GGTTGCCGCA CTCGCCAACC TAGCCGTGCC GCGCGGCT CGCGCCGCGG GCGCGCTA 960 TTGGCACACG AGCGGCCGGG AGATCCTGGA CGCGAACAAC GTGCCGGTAC GGATCGCCGG 1020 CATCAACTGG TTTGGGTTCG AAACCTGCAA TTACGTCGTG CACGGTCTCT GGTCACGCGA 1080 CTACCGCAGC ATGCTCGACC AGATAAAGTC GCCGAACAAC GTGCCGGTAC GGATCGCCGA 1140 CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAGC ATCAATTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG 1260 TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCACGC GCACTCGCC 1320 CCTGTGGTAC ACGACCACC TCTCGGAGCC TACGTGGAT TCCGACCTGC AAACCACATCC CGTCGCCTA 1380 GCAGCGCTAC AAGGGAAACC CGACCGTCT CGGCTTTGAC TTCCGACCTGC AAGCGCTGGC 1380 CCAGGGCTTC AGGGGACCC CGACCGTCGT CGGCTTTGAC TTCCGACCACC AGCGCTGCC 1380 CCAGGCCTGC TGGGGCTCCG GCGATCCGAC CATCGACTG CGATTGGCCG CCGACCGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTACACACG AGCCGCATGA 1440 CCCGCAGAACGC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTACACACG AGCCGCATGA 1560 CTACAAACGGA GACTCCTACT GGTGGGGGCG CAACCTGCAA GGACCGGCC AGTACCCCGGT 1620 CGGGCTGAAC GCCGTCACC GCCTGGTGTA CTCGGCGCA GACCGCCCA GAACCCGCA 1660 CCCGCAGACG TGGTTCAGCG ACCAGACCTT CCCCAACACA ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGAA TACCTCTCA ATCAGAACAT TCCACCGGTA TGGCCGGCA TCTGGAACAA 1740 GAACTGCAA TCCACGACCG ACCAGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGCAA TACCGTCCGA CCAGACCTT CCCCAACACA TCGCCACGCA TACCCGGT 1800 GACACTGCAA TACCGTCCG ACCAGACCTT CCCCAACAC TCGCAC TCTGCACACA 1740 GACCGCCACAC GGAGAACT TCAAGGACCT CCCGACCTT TCGGCCC AACCCCGCA TCTGCACCC 1860 GCCGCACACC CCGACCAG TCCAGGCC ACCGCCC TCCTCCAGT ACCTCCCACCC 1860 GCCGACCCC CCGCCCGC TCCCGCCC TCCGCCC ACCGCC CTCCTCCGAC CCCCCCCACGCC CCCCCCCCCC	CGCAGAGTAG	GCGACTGTAT	GCGGTAGGTT	GGCGCTCCAG	CCGTGGGCTG	GACATGCCTG	660
ACGGGGGAGA ACAGACGGGG GAGAAACCAA CGGGGGATTG GCGGTGCCGC GCGCATTGCC GCGAGTGCCT GGCTCGCGGG TGATGCTGCG GGTCGGCGT CGCGCCGCG TGCTGGCATT 900 GGTTGCCGCA CTCGCCAACC TAGCCGTGCC GCGGCGGCT CGCGCCGCGG GCGGCGGCTA 960 TTGGCACACG AGCGGCCGGG AGATCCTGGA CGCGAACAAC GTGCCGGTAC GGATCGCCGG 1020 CATCAACTGG TTTGGGTTCG AAACCTGCAA TTACGTCGTG CACGGTCTCT GGTCACGCGA 1080 CTACCGCAGC ATGCTCGACC AGATAAAGTC GCTCGGCTAC AACACAATCC C7CTGCCGTA 1140 CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAAC ATCAATTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAAATCGTCG CGTACGCCGG 1260 TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACGG GATTGCAGGG GGCAGTCGGC 1320 GCTGTGGTAC ACGAGCAGC TCTCGGAGGC TACGTGGAT TCCGACCTG AAAAATCGTCG CGTACGCCGG 1320 GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTGCACAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTCGG GCGATCCGAC CATCGACTGG CGATTGGCCG CGGACGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTATT TTCGTACACACG AGCCGCATGA 1440 CCCGCAGACG GACTCCTACT GGTGGGGGGC CAACCTGCAA GGACCGCGCC AGTACCCCGGT 1620 CTACAACGGA GACTCCTACT GGTGGGGGGG CAACCTGCAA GGACCGCGCC AGTACCCCGGT 1620 CGGCGGAAC GCCGTCACT GGTGGGGGGG CAACCTGCAA GGACCGCGC AGTACCCGGT 1620 CCGCCAGACG TGGTTCAGCG ACCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA CCCGCCAGACG TGGTTCAGCG ACCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA CCCGCAGACG TGGTTCACGA CCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA CACACTGCAA TCCACGACCT ACCAGACCTG CCCGACACAC ATCCCCGGCA TCTGGAACAA CACACTGCAA TCCACGACCT ACCAGACCTG CCCGACACAC ATCCCCGGCA TCTGGAACAA CACACTGCAA TCCACGACCT ACCAGACCTG CCCGACGCT TCGGACCACA TAAAAGACGG CACACTGCAA TCCACGACCT ACCAGACCTC TCGAACACA TTGCCCGGCA AATTCGGTAC 1800 CACACTGCAA TCCACGACCT ACCAGACCT TCCACACAC TACACGCC 1860 CACACTGCAA TCCACGACCT TCCAGACCTC TCGACACAG TAAAAAGACGG 1980 CTATCTCCCC CCGACGCC ACCAGCCGCA CCGCCCACACAC TAAAAGACGG 1980 CTATCTCCCC CCGACGCCC TCCGCCCC ACCGCCC CCCCCCCGACCC CCCCCCCACGCC CCCCCCCC	CTGCGAACTC	TTGACACGTC	TGGTTGAACG	CGCAATACTC	CCAACACCGA	TGGGATCGTT	720
GCGAGTGCCT GGCTCGCGGG TGATGCTGCG GGTCGGCGTC GTCGTCGCGG TGCTGGCATT 900 GGTTGCCGCA CTCGCCAACC TAGCCGTGCC GCGGCGGCT CGCGCCGCGG GCGCGGCTA 960 TTGGCACACG AGCGGCCGGG AGATCCTGCA CGCGACAAC GTGCCGGTAC GGATCGCCGG 1020 CATCAACTGG TTTGGGTTCG AAACCTGCAA TTACGTCGTG CACGGTTCT GGTCACGCGA 1080 CTACCGCAGC ATGCTCGACC AGATAAAGTC GCTCGGCTAC AACACAATCC GGCTGCCGTA 1140 CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAAC ATCAATTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG 1260 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG 1260 TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCAGCG GGCAGTCCGC 1320 GCTGTGGGTAC ACGAGCAGC TCTCGGAGGC TACGTGGATT TCCGACCTG AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACCGTCT CGGCTTTGAC TTGCACAACG AGCCGCTGGC 1380 GCAGCGCTGC TGGGGCTGC GCGATCCGAG CATCGACTG CGATTGGCCG CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAC AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGC GCGATCCGAC CACCGACCG CGATCGACG CGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAC GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTTA CTCGGCGCA GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCT CCCCAACAAC ATGCCCGGC AGTCCCGGT 1620 GACACTGCAA TCCACACCG ACCAGACCT CCCCAACAAC ATGCCCGGCA ACTTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACCTC GCTGAACAC ATGCCCGGCA ACTCCGGTA 1620 CGCGCACACA GCAGGAATTC TCAAGGATGA CTGGAACAG TACAGCGC AACCCGGATC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGAG GTCGACACA TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGGATTC TCGGCAGACG GTCGACACA TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGGATTC TCGGCAGAC GTCGACCAC TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGCTTC TCCGTCGCC AGCCCTACG CAACCCGGATC 1920 CGCGACGCCA ACCCCGCTG TCTCGCCG ACCCGCCC AGCCCCAACC CCGACCCC ACCCCCACACC CCGACCCC ACGCCCAACC CCGCCCCACCCC CCGCCCCCCCC	CCCATAAGTT	TCCGTCTCAC	AACAGAATCG	GTGCGCCCTC	ATGATCAACG	TGAAAGGAGT	780
GGTTGCCGCA CTCGCCACC TAGCCGTGCC GCGGCCGGCT CGCGCCGCG GCGGCGCTA 960 TTGGCACACG AGCGGCCGG AGATCCTGGA CGCGAACAAC GTGCCGGTAC GGATCGCCGG 1020 CATCAACTGG TTTGGGTTCC AAACCTGCAA TTACGTCGTG CACGGTCTCT GGTCACGCGA 1080 CTACCGCAGC ATGCTCGACC AGATAAAGTC GCCGAACAAC ACCACAATCC GGCTGCCGTA 1140 CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAGC ATCAATTTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG 1260 TCAGGACCGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCAGCG GGCAGTCGGC 1320 GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTCCACAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGC GCGATCCGAG CATCGACTGG CGATTGGCC CCGAGCGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAC AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGC GCGATCCGAA CCTGCTCATT TTCGTCGAAC GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGACGCGCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCCGGCA TCTGGAACAA 1740 GAACTGCGAA TCCACGACCG ACCAGACGTG GCTGAAGACA TCCGCGCA ACCTGCAA ACCCCGGCT 1860 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACA TCCGACCGC 1860 GACCCGCGCAA TACGGTCGG ACCAGACGTG GCTGAAGACA TCGGCCGC 1860 GACCCGCGCAA TACGGTCGG ACAGCTTCCA GTGGACCTTC TGGTCCAGT ACCTACGGCC 1860 CACCGCGCAA TACGGTCGG ACAGCTTCCA GTGGACCTTC TGGTCCAGT ACCTACGGCC 1860 CACCGCGCAA TACGGTCGG ACAGCTTCCA GTGGACCTTC TGGTCCAGT ACCTACGGCC 1860 CACCGCGCAA TACGGTCGG ACAGCTTCCA GTGGACCTTC TGGTCCAGT ACCTACGGC 1980 CAGCCACCAC GGAAGAATTC TCAAGGATGA CTGGCACAG TAAAAGACG 1980 CTATCTCGCC CCGATCAAGT CTCAAGGATGA CTGCCGCG ACCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGCCGCG ACCCGCTCGC CAGGCCCA CCGATCCAG 1980 CTATCTCGCC CCGATCAAGT CTCCAGCGCC TCCGCCGC ACCCCGCCC ACCCCCACGCC ACCCCGACCC CTACTGCTAC 2160 CCCACGCCC ACGCCACACC CCGACGCCA ACCCGCAGCCC ACCCCGCGCC CTACTGCCC CTACTGCTAC 2220 CCCACGCCC ACGCCAACCC CCAACGCCA CCCGACGCCA ACCCGTGCC CCCCGTGCAC CCCGCTGCAC CCGCTGCAC CCGC	ACGGGGGAGA	ACAGACGGGG	GAGAAACCAA	CGGGGGATTG	GCGGTGCCGC	GCGCATTGCG	840
TTGGCACACG AGCGGCCGGG AGATCCTGGA CGCGAACAAC GTGCCGGTAC GGATCGCCGG 1020 CATCAACTGG TTTGGGTTCG AAACCTGCAA TTACGTCGTG CACGGTCTCT GGTCACGCGA 1080 CTACCGCAGC ATGCTCGACC AGATAAAGTC GCTCGGCTAC AACACAATCC GCTCGCCGTA 1140 CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACACG ATCAATTTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC ATCAATTTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG 1260 TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCAGCG GGCAGTCGGC 1320 GCTGTGGTAC ACGACGACG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGGCCTAC AAGGGAACC CGACGGTCGT CGGCTTTGAC TTGCACAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGCG GCGATCCGAC CATCGACTG CGATTGGCCG CCGACGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAC GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCG CAACCTGCAA GGAGCCGGC AGTACCCGGT 1620 CGTGCTGAAC GTGCTCACG GCCTGGTGTA CTCGGCCGA CGACCTGCAA GGAGCCGGC AGTACCCGGT 1620 CGTGCTGAAC GTGCTCAGC GCCTGGTGTA CTCGGCGCA CACTACCACA AGCCCGGTC 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCCGA CACTACCGGA CGAGCCTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGCCG AATTCCGGTAC 1800 GACACTGCAA TCCAGACCG ACCAGACGT GCTGAAGACG TCTGGAACAA 1740 GAACTGGGA TACCTCTTCA ATCAGACAT TGCACCGGTA TCGGCCGC AATTCCGGCC 1860 GACCGCCACAA TACGGTCCGG ACCAGACGT GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1800 GACCGCGCAA TACGGTCCGG ACCAGACGT GCTGAAGACG TCCGTCCAGT 1920 CGGCGAACAC GGAGGAATTC TCAAGGATGA CTGGCAGACG TCGTCCAGT ACCTACGGCC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGCC TCGTCCAGT ACCTACGCC 1920 CGGCGACACC GCGACACAC CCGCCGCC ACGCCCA ACCCCGATC 1920 CAGTCAACCG TCCCCGTCGG TCTCGCCGC ACCCCGCC ACCCCGCCC ACCCCCACCC ACCCCCACCC ACCCCCACCC ACCCCCTCGC CCACCCCACC	GCGAGTGCCT	GGCTCGCGGG	TGATGCTGCG	GGTCGGCGTC	GTCGTCGCGG	TGCTGGCATT	900
CATCAACTGG TTTGGGTTCG AAACCTGCAA TTACGTCGTG CACGGTCTCT GGTCACGCGA 1080 CTACCGCAGC ATGCTCGACC AGATAAAGTC GCTCGGCTAC AACACAATCC GCTGCCGTA 1140 CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAGC ATCAATTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG 1260 TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCACGC GGCAGTCGGC 1320 GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACCGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTCCGACACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTCGC GCGATCCGAC CATCGACTG CGATTGGCCC CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTATT TTCGTCGAAG GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCGG CAACCTGCAA GGAGCCGCCA AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC AGCCGCACCA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGA TACCTCTCA ATCAGAACAT TGCACCGGTA TGGCTGGCG AATTCGGTAC 1800 GACACTGCAA TCCACACCG ACCAGACCTG CTGGAAGAC CTCGCTCAGT ACCTACGCC 1860 GACCCGCAA TACCGGTGCG ACACCTTCCA GTGGACCATC TGCCCGACACA ACCCCGGTA 1800 GACACTGCAA TCCACACCC ACCAGACCTG CTGGAACAC TTGGCTCGGC AATTCCGGTAC 1820 GACCGCGCAA TACCGGTGCG ACACCTTCCA GTGGACCTTC TGGCTCAGTA ACCTACGGCC 1860 GACCGCGCAA TACCGTGCGG ACAGCTTCCA GTGGACCTTC TGGCTCAGTA ACCTACGGCC 1860 CTATCTCGCG CCGATCAAGT CTCAAGGATGA CTGGCAGACG TAAAAAACAGG 1980 CTATCTCGCG CCGATCAAGT CTCAAGGATGA CTGGCAGACAG TAAAAAACAGG 1980 CTATCTCCGC CCGATCAAGT CTCAAGGATGA CTGGCAGACAG TAAAAAACAGG 1980 CTATCTCCGC CCGATCAAGT CTCCAGCTC TCCGTCCGC AGCCCTCGG CAGTCCGAC CAGGCCCC ACCGCCCG TCCCGCCGC ACCCGCCCA ACCCCGCCC CTACTGCCAC CCCACGCCC ACGCCAACCC CCGACGCCA ACCCGCCCA ACCCTGCCC CTACTGCCAC CCCACGCCC ACGCCAACCA GCACGCCAA CCCGCCCCAACCAC CTACTGCCCAC 2220 CCCCACGCCC ACGGCAACCA GCACGCCAA CCCGCCCCAACCGCCA ACCCCGCCCA CCCCCCCC	GGTTGCCGCA	CTCGCCAACC	TAGCCGTGCC	GCGGCCGGCT	CGCGCCGCGG	GCGGCGGCTA	960
CTACCGCAGC ATGCTCGACC AGATAAAGTC GCTCGGCTAC AACACAATCC GGCTGCCGTA CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAGC ATCAATTTT ACCAGATGAA TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAAATCGTCG CGTACGCCGG TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCAGCG GGCAGTCGGC GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC CAGCGCTAC AAGGGAAACC CGACCGTCGT CGGCTTTGAC TTGCACAACG AGCCGCATGA CCCGGCCTGC TGGGGCTGCG GCGATCCGAG CATCGACTGG CGATTGGCCG CCGACGGGC CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAG GTGTGCAGAG CTACAACGGA GACTCCTACT GGTGGGGCG CAACCTGCAA GGAGCCGCC AGTACCCGGT CCGGCAGACC GTGCTCACT GCTGGGGCG CAACCTGCAA GGAGCCGCC AGTACCCGGT CCCGCAGACC TGGTTCAGC ATCCGACCT CCCCAACAAC ATGCCCGGC AGTACCCGGT CCCGCAGACC TGGTTCAGCG ATCCGACCT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTCA ATCAGAACAT TGCACCGGTA TGGCTGGCC AATTCCGTAC GACCTGCAA TCCACGACCG ACCAGACCT GCTGAAGACC CTCGTCCAGT ACCTACGGCC 1860 GACCGCCAAA TACGGTGCGG ACAGACCTT CGGCGCAC CTCGTCCAGT ACCTACGGCC 1860 GACCGCCCAA TACGGTGCGG ACAGCTTCCA GTGGAACAC CTCGTCCAGT ACCTACGGCC 1860 CCGGCGCACACA GGAGGAATTC TCAAGGATGA CTGGCAGACC GTCGACCACAG TAAAAGACGG 1980 CTATCTCGCC CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCCAGT CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCCGCC AGCCCTCAG CACCCCTACC 2100 CCCGACGCCC ACGCACACC CGACAGCCAG CCCGACCCC ACCCCGACCC CTACTGCCAC 2220 CCCCACGCCC ACGGCAACA GCGATCGGGC CAACGCCGCA CCCCACGCCC CTACTGCCAC 2220 CCCCACGCCC ACGGCAACA GCGATTGGGG CAATGGGCTC ACCGCCTGAC CCCGCCTGCAC 2220	TTGGCACACG	AGCGGCCGGG	AGATCCTGGA	CGCGAACAAC	GTGCCGGTAC	GGATCGCCGG	1020
TCTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAGC ATCAATTTTT ACCAGATGAA TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCAGCG GGCAGTCGGC 1320 GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTGCACAAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGCG GCGATCCGAG CATCGACTGG CGATTGGCCG CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGGTCATT TTCGTCGAAG GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCGG CAACCTGCAA GGAGCCGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGAGCG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTCA ATCAGAACAT TGCACCGGTA TGGCTGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCCCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 GGCGGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCC CCGATCAAGT CGTCGATTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 QAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCCCGGA CCGAGTCGGAC 2100 QAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCCCGA CCCCCTTCCAC 2160 GCCGACGCCA ACTCCGACCC CGACAGCCAG CCCGACGCCA ACGCTGACC CTACTGCTAC 2220 CGCGAGTTAC CAGGCCC CGACAGCCGT ACCGACGCA ACGCTGACCC CTACTGCTAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGGCT ACCGCCTGAC 2220	CATCAACTGG	TTTGGGTTCG	AAACCTGCAA	TTACGTCGTG	CACGGTCTCT	GGTCACGCGA	1080
TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG 1260 TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCAGCG GGCAGTCGGC 1320 GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTGCACAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGCG GCGATCCGAG CATCGACTGG CGATTGGCCG CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAG GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCGG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGT GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 QAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC 2100 GCCGACGCCC ACCGCACGC CGACAGCCAG CCCGACGCC ACCGCTCGC CAACGCCCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGC CAACGGCTA ACCGCTGACC CTACTGCTAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGC CAACGGCTA ACCGCTGGAC CCCGCTGCAC 2220	CTACCGCAGC	ATGCTCGACC	AGATAAAGTC	GCTCGGCTAC	AACACAATCC	GOCTGCCGTA	1140
TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCAGCG GGCAGTCGGC 1320 GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTGCACAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGCG GCGATCCGAG CATCGACTGG CGATTGGCCG CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAG GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACCG GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGC ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTCA ATCAGAACAT TGCACCGGTA TGGCTGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACCT GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTCGGA ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1820 CGACCGCGAA TACGGTGCGG ACACGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGACCAA ACCCCGTTAC 2040 CAGTCAACCC TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC 2100 GCCGACGCCC ACGCCCAACACC CCGACGCCA ACGCTGACC CTACTGCTAC 2160 GCCCACGCCC ACGCCCAACGCC CGACGCCAA ACGCTGACCC CTACTGCTAC 2220 CCCCACGCCC ACGGCAACAA GCGATTGGGG CAATGGGCTC ACCGACGCCA ACGCTGACC CTACTGCTAC 2220 CCCCACGCCC ACGGCAACAA GCGCTTACAACG TGGCCGTGAC CCCGCTGACC CCACGCCCA ACGCCTGACC CCACGCCCA ACGCCTGACC CCACGCCCA ACGCCTGACC CTACTGCTAC 2220 CCCCACGCCC ACGGCAACAA GCGATTGGGG CCAATGGCTTA ACCGGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGACC CTACTGCTAC 2220 CCCCACGCCC ACGGCAACAA GCGCTGACC CCAATGGCTTA ACCGGTAACGG TGGCCGTGAC CCCGCTGAC CCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCAGCCCA ACGCTTACCGCTTAC CCGCTGAC CCCGCTGAC CCCGCTGAC CCCGCTCGAC CCCGCTCACCCC CTACTGCTAC CCGCTTAC CCGC	CTCTGACGAC	ATTCTCAAGC	CGGGCACCAT	GCCGAACAGC	ATCAATTTTT	ACCAGATGAA	1200
GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTGCACAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGCG GCGATCCGAG CATCGACTGG CGATTGGCCG CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAG GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCGG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCCCG AGCCCGTCGAC 2220 CCCGACGCCC ACGGCAAGCC CGACAGCCAG CCCGACCCC ACGCCCCACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACAGCCAG CCCGACGCC ACGCTGGAC CCTACTGCTAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACCGTGACC CTACTGCTAC 2220 CCCCACGCCC ACGGCAAGCC CGACAGCCGC CAATGGCTTC ACCGTGACC CTACTGCTAC 2220	TCAGGACCTG	CAGGGTCTGA	CGTCCTTGCA	GGTCATGGAC	AAAATCGTCG	CGTACGCCGG	1260
GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTGCACAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGCG GCGATCCGAG CATCGACTGG CGATTGGCCG CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAG GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCGG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 QAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCG AGCCCGTCGGAC 2100 GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAACA GCGATGCGGC CCCGACGCCA ACCGCTGACC CTACTGCTAC 2220 CGCGAGTTAC CAGGCAACA GCGATTGGGG CAATGGCTTC ACGGCTGAC CCCGATGCC 2220	TCAGATCGGC	CTGCGCATCA	TTCTTGACCG	CCACCGACCG	GATTGCAGCG	GGCAGTCGGC	1320
CCCGGCCTGC TGGGGCTGCG GCGATCCGAG CATCGACTGG CGATTGGCCG CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAG GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCGG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCC AGCCCGTCGG CGAGTCGGAC 2100 GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGCCA ACGCTGACCC CTACTGCTAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGC CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2220	GCTGTGGTAC	ACGAGCAGCG	TCTCGGAGGC	TACGTGGATT	TCCGACCTGC	AAGCGCTGGC	1380
CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAG GTGTGCAGAG CTACAACGGA GACTCCTACT GGTGGGGCGG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCC AGCCCGTCGG CGAGTCGGAC 2100 GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGGCA GCCTCCGGAG CCCGCTGCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2220	GCAGCGCTAC	AAGGGAAACC	CGACGGTCGT	CGGCTTTGAC	TTGCACAACG	AGCCGCATGA	1440
CTACAACGGA GACTCCTACT GGTGGGGCG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGGAC CGAGTCGGAC 2100 GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGCCA GCCTCCGGAG CCCGCTGCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2220	CCCGGCCTGC	TGGGGCTGCG	GCGATCCGAG	CATCGACTGG	CGATTGGCCG	CCGAGCGGGC	1500
CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC 2100 GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGCA GCCTCCGGAG CCCGCTGCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	CGGAAACGCC	GTGCTCTCGG	TGAATCCGAA	CCTGCTCATT	TTCGTCGAAG	GTGTGCAGAG	1560
CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGGCG AATTCGGTAC 1800 GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC 1860 GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC 2100 GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGCA GCCTCCGGAG CCCGCTGCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	CTACAACGGA	GACTCCTACT	GGTGGGGCGG	CAACCTGCAA	GGAGCCGGCC	AGTACCCGGT	1620
GAACTGGGGA TACCTCTCA ATCAGAACAT TGCACCGGTA TGGCTGGGCG AATTCGGTAC GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGGCA GCCTCCGGAG CCCGCTGCAC CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC CCGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	CGTGCTGAAC	GTGCCGAACC	GCCTGGTGTA	CTCGGCGCAC	GACTACGCGA	CGAGCGTCTA	1680
GACACTGCAA TCCACGACCG ACCAGACGTG GCTGAAGACG CTCGTCCAGT ACCTACGGCC GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC 1920 CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGGCA GCCTCCGGAG CCCGCTGCAC CGCGAGGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	CCCGCAGACG	TGGTTCAGCG	ATCCGACCTT	CCCCAACAAC	ATGCCCGGCA	TCTGGAACAA	1740
GACCGCGCAA TACGGTGCGG ACAGCTTCCA GTGGACCTTC TGGTCCTGGA ACCCCGATTC CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG QAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGGCA GCCTCCGGAG CCCGCTGCAC CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	GAACTGGGGA	TACCTCTTCA	ATCAGAACAT	TGCACCGGTA	TGGCTGGGCG	AATTCGGTAC	1800
CGGCGACACA GGAGGAATTC TCAAGGATGA CTGGCAGACG GTCGACACAG TAAAAGACGG 1980 CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC 2100 GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGGCA GCCTCCGGAG CCCGCTGCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	GACACTGCAA	TCCACGACCG	ACCAGACGTG	GCTGAAGACG	CTCGTCCAGT	ACCTACGGCC	1860
CTATCTCGCG CCGATCAAGT CGTCGATTTT CGATCCTGTC TAATGATCTG CATCGCCTAG 2040 CAGTCAACCG TCCCCGTCGG TGTCGCCGTC TCCGTCGCCG AGCCCGTCGG CGAGTCGGAC 2100 GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGGCA GCCTCCGGAG CCCGCTGCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	GACCGCGCAA	TACGGTGCGG	ACAGCTTCCA	GTGGACCTTC	TGGTCCTGGA	ACCCCGATTC	1920
CAGTCAACCGTCCCCGTCGGTGTCGCCGTCTCCGTCGCCGAGCCCGTCGGCGAGTCGGACGCCGACGCCTACTCCGACGCCGACAGCCAGCCCGACGCCAACGCTGACCCCTACTGCTAC2160GCCCACGCCCACGGCAAGCCCGACGCCGTCACCGACGCAGGCCTCCGGAGCCCGCTGCAC2220CGCGAGTTACCAGGTCAACAGCGATTGGGGCAATGGCTTCACGGTAACGGTGGCCGTGAC2280	CGGCGACACA	GGAGGAATTC	TCAAGGATGA	CTGGCAGACG	GTCGACACAG	TAAAAGACGG	1980
GCCGACGCCT ACTCCGACGC CGACAGCCAG CCCGACGCCA ACGCTGACCC CTACTGCTAC 2160 GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGGCA GCCTCCGGAG CCCGCTGCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	CTATCTCGCG	CCGATCAAGT	CGTCGATTTT	CGATCCTGTC	TAATGATCTG	CATCGCCTAG	2040
GCCCACGCCC ACGGCAAGCC CGACGCCGTC ACCGACGGCA GCCTCCGGAG CCCGCTGCAC 2220 CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	CAGTCAACCG	TCCCCGTCGG	TGTCGCCGTC	TCCGTCGCCG	AGCCCGTCGG	CGAGTCGGAC	2100
CGCGAGTTAC CAGGTCAACA GCGATTGGGG CAATGGCTTC ACGGTAACGG TGGCCGTGAC 2280	GCCGACGCCT	ACTCCGACGC	CGACAGCCAG	CCCGACGCCA	ACGCTGACCC	CTACTGCTAC	2160
	GCCCACGCCC	ACGGCAAGCC	CGACGCCGTC	ACCGACGGCA	GCCTCCGGAG	CCCGCTGCAC	2220
AAATTCCGGA TCC 2293	CGCGAGTTAC	CAGGTCAACA	GCGATTGGGG	CAATGGCTTC	ACGGTAACGG	TGGCCGTGAC	2280
	AAATTCCGGA	TCC					2293

Fig. 6

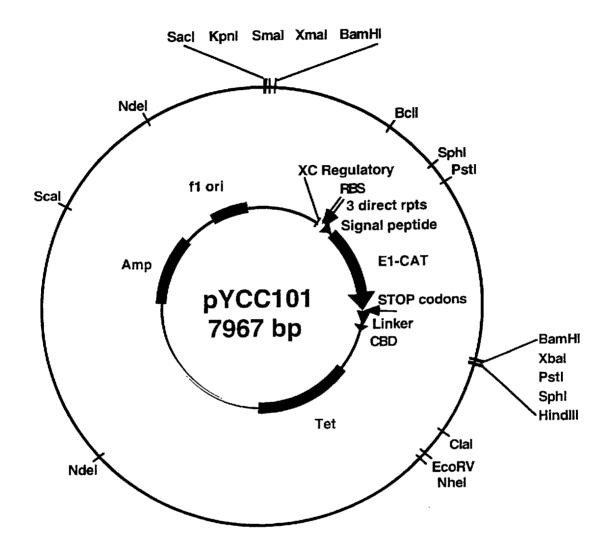
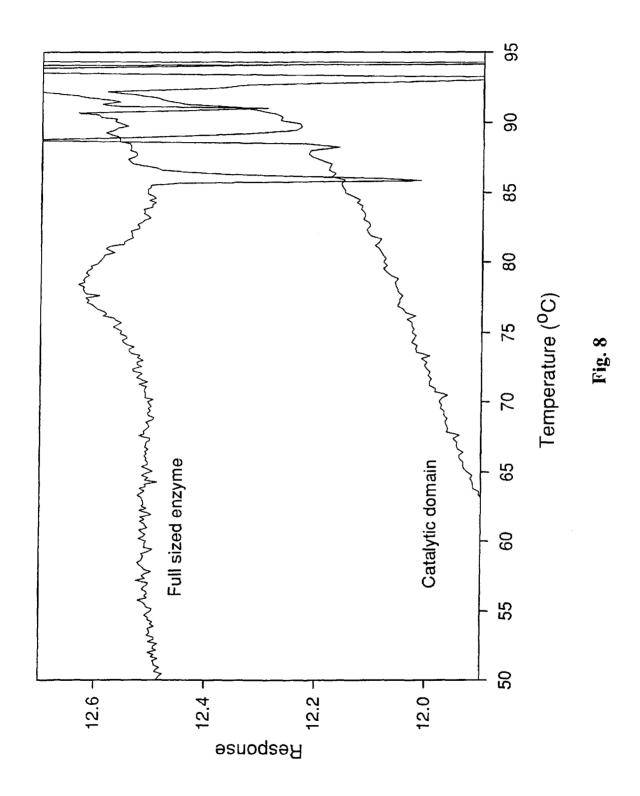


Fig. 7



METHOD FOR INCREASING THERMOSTABILITY IN CELLULASE ENNZYMES

This application is a continuation-in-part of Ser. No. 5 08/276,213, filed Jul. 15, 1994, now U.S. Pat. No. 5,536, 655, which is a continuation-in-part of Ser. No. 08/125,115, filed Sep. 21, 1993, now U.S. Pat. No. 5,366,884, which is a continuation-in-part of Ser. No. 07/826,089, filed Jan. 27, 1992, now U.S. Pat. No. 5,275,944, which was a 10 continuation-in-part of Ser. No. 412,434, filed Sep. 26, 1989 now U.S. Pat. No. 5,110,735.

The United States Government has rights in this invention under Contract No. DE-AC36-83CH10093 between the United States Department of Energy and the National 15 Renewable Energy Laboratory, a Division of the Midwest Research Institute.

FIELD OF THE INVENTION

The invention relates to methods for increasing the thermostability of cellulose enzymes.

BACKGROUND OF THE INVENTION

The development of an economic process for the conversion of low-value biomass to useful products via fermentation requires the optimization of several key steps, including cellulase production and performance. Practical utilization of cellulose by hydrolysis with cellulase to produce glucose requires large amounts of cellulase to fully depolymerize 30 cellulose. For example, about one kilogram cellulase preparation may be used for every fifty kilograms of cellulose. Economical production of cellulose is also compounded by the relatively slow growth rates of cellulase-producing fungi and the long times required for cellulose induction. 35 Therefore, improvements in or alternative cellulase production systems capable of greater productivities of cellulase activity than may be possible from currently available systems would significantly reduce the cost of cellulose hydrolysis and make the large-scale bioconversion of cellulosic biomass more economical.

Highly thermostable cellulose enzymes are secreted by the cellulolytic thermophile Acidothermus cellulolyticus gen. nov., sp. nov. These are discussed in U.S. Pat. Nos. 5.110.735, 5.275.944, 5.366,884, and 5,432,075. All four of these patents are included herein in their entirety by this reference thereto. This bacterium was originally isolated from decaying wood in an acidic, thermal pool at Yellowstone National Park and deposited with the American Type Culture Collection (ATCC) under collection number 43068 (Mohagheghi et al. 1986. Int. J. System. Bacteriol. 36:435-443).

The cellulase complex produced by this organism is known to contain several different cellulase enzymes with maximal activities at temperatures of 75° C. to 83° C. These 55 cellulases are resistant to inhibition from cellobiose, an end product of the reactions catalyzed by cellulase. Also, the cellulases from Acidothermus cellulolyticus are active over a broad pH range centered about pH 6. A high molecular weight cellulase isolated from growth broths of Acidothermus cellulolyticus was found to have a molecular weight of approximately 156,600 to 203,400 daltons by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). This enzyme is described in U.S. Pat. No. 5,110,735.

A novel cellulase enzyme, known as the E1 65 endoglucanase, also secreted by *Acidothermus cellulolyticus* into the growth medium, is described in detail in U.S. Pat.

2

No. 5.275,944. In its native form, this endoglucanase demonstrates a temperature optimum of 83° C. and a specific activity of 40 µmole glucose release from carboxymethylcellulose/min/mg protein. This E1 endoglucanase was further identified as having an isoelectric pH of 6.7. It is this E1 endoglucanase which has been modified and made the subject of this patent application. The E1 endoglucanase is a multidomain cellulase having a catalytic domain and a cellulose binding domain connected to the catalytic domain by a linker peptide.

SUMMARY OF THE INVENTION

The E1 endoglucanase described above, has been modified to increase its thermostability. The present modification has increased the thermostability of this enzyme by effectively doubling the length of time which this enzyme demonstrates half-maximal activity at elevated temperatures, as well as increasing the temperature at which maximal rates of catalysis are observed. The modification comprises eliminating the cellulose binding domain and linker peptide of the enzyme from the catalytic domain. It is the catalytic domain containing the catalytically active portion of the molecule. which remains after elimination of the cellulose binding domain and linker peptide, which demonstrates these improved thermal properties. This modification has been accomplished by two methods. The first method for eliminating the cellulose binding domain and linker peptide is by subjecting the entire molecule to proteolytic cleavage, which removes the cellulose binding domain and linker peptide from the catalytic domain. The second method for removing the cellulose binding domain and linker peptide from full size E1 involves modification of the gene which encodes full size E1 so that the cellulose binding domain and linker peptide are not present in the expression product. The E1 enzyme of the present invention which demonstrates enhanced thermostability by elimination of the cellulose binding domain and linker peptide is referred to as "modified", "truncated", "catalytic domain", or "E1 CAT". Also, E1 CAT produced by proteolytic cleavage may be referred to as pE1 CAT and E1 CAT produced by genetic transformation may be referred to as gel CAT.

In addition to the modified E1 endoglucanase having improved thermostability, this invention teaches the expression of full size E1 in a yeast

It is therefore an object of the present invention to transform and express the full size E1 endoglucanase gene in a yeast under the same and/or a different gene regulatory system.

It is an object of the present invention to modify the gene encoding the E1 endoglucanase from *Acidothermus cellulolyticus* to enhance its thermostability by eliminating expression of the cellulose binding domain and linker peptide in the gene product.

It is an object of the present invention to provide the DNA sequence which encodes the modified form of the E1 endoglucanase from Acidothermus cellulolyticus.

It is another object of the present invention to provide the amino acid sequence of the modified form of the E1 endoglucanase from Acidothermus cellulolyticus.

It is another object of the present invention to provide a method for proteolytic cleavage of the cellulose binding domain and linker peptide from the E1 endoglucanase.

It is a further object of the present invention to prepare modified E1 endoglucanases which have different properties from the natural enzyme.

The present invention describes the gene for and the nucleotide sequence of the segment of Acidothermus cellu3,712,1

lolyticus DNA encoding the catalytic domain of the E1 endoglucanase gene. This 2293 base fragment of DNA is unique in nature and discretely defined. The natural gene contains a promoter, a ribosome binding site, a signal peptide, an open reading frame, a termination codon and a 5 putative transcriptional terminator. The modified gene contains a promoter, a ribosome binding site, a signal peptide and one or more termination codons inserted at the C terminus of the catalytic domain.

3

The cloned gene may also be expressed in other microorganisms under its natural promotor or another promotor recognized by the host microorganism. The cloned gene may be expressed in mammalian systems, higher plants or vital vectors. Alternatively, additional copies of the gene may be introduced into Acidothermus cellulolyticus or other heterologous host organism to enhance expression of the enzyme. Additionally, DNA encoding one or more domains of the Acidothermus cellulolyticus E1 endoglucanase may be ligated to domains in other compatible endoglucanases to make a recombinant DNA capable of expressing a hybrid endoglucanase enzyme having beneficial properties from both endoglucanases.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a plasmid map of pPIC9-E1.

FIG. 2 compares the functional half-life of the full size E1 enzyme to the functional half-life of the catalytic domain of the E1 enzyme at 80° C.

FIG. 3 compares the temperature optimum of the full size $_{30}$ E1 enzyme to the temperature optimum of the catalytic domain of the E1 enzyme.

FIG. 4 is a chromatogram showing purified constituent domains of full size E1 by size exclusion chromatography.

FIG. 5 shows the amino acid translation of the coding 35 tomyces lividans. sequence of E1 CAT. Expressing full

FIG. 6 shows the 2293 base pair nucleotide sequence of the region of Acidothermus cellulolyticus genomic DNA which contains the modified E1 endoglucanase gene which expresses only the catalytic domain of the enzyme, without the linker peptide or cellulose binding domain.

FIG. 7 is a plasmid map of pYCC101.

FIG. 8 is a DSC thermogram comparing the denaturation endotherm peak at 78° C. for the full size E1 and E1 CAT $_{45}$ showing no peak below at least 88° C.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

According to the present invention the entire E1 coding sequence for Acidothermus cellulolyticus E1 endoglucanase is cloned and expressed in a different microbial host than is described in the parent to this application, Ser. No. 08/276, 213, filed Jul. 15, 1994, which is incorporated herein in its entirety by this reference thereto. The enzyme described in Ser. No. 08/276,213 is a β -1,4 endoglucanase which can hydrolyze cellulose or carboxymethylcellulose and is hereafter referred to as E1 endoglucanase or full size E1. The result is a vastly improved rate of E1 enzyme production over native or genetically engineered bacterial systems, 60 thereby lowering the cost of cellulase. The instant application teaches expression of full size E1 in a yeast, namely Pichia pastoris.

Expression of Full Size E1 and Modified E1

For expressing the E1 endoglucanase gene, either the full size E1 gene or the modified E1 gene, one may use a variety

of hosts including most bacteria, yeast, fungi, algae, viruses, plants and animals. Organisms which naturally produce cellulase enzymes are preferred host cells along with easy to grow host cells and host cells known to express and secrete heterologous genes in large quantities.

If the host cell is a bacterium, generally a bacterial promoter and regulatory system will be used. For a typical bacterium such as $E.\ coli$, representative examples of well known promoters include, for example, trc, lac, tac, trp, bacteriophage lambda P_L , T7 RNA polymerase promoter, etc. When the expression system is yeast, examples of well known promoters include, but are not limited to GAL 1/GAL 10, alcohol dehydrogenase (ADH), alcohol oxidase (AOX), his3, cycI, etc. For eukaryotic hosts, enhancers such as the yeast Ty enhancer, may be used.

Alternatively, if one wished for the full size or modified E1 endoglucanase gene to be expressed at only a particular time, such as after the culture or host organism has reached maturity, an externally regulated, environmentally-responsive promoter is particularly useful. Examples include those based upon the nutritional or chemical composition of the medium (e.g. methanol, lac, trp, his), temperature regulation (e.g. temperature sensitive regulatory elements), heat shock promoters (e.g. HSP80A, U.S. Pat. No. 5,187,267), stress response (e.g. plant EF1A promoter, U.S. Pat. No. 5,177,011) and chemically inducible promoters (e.g. tetracycline inducible promoter or salicylate inducible promoter U.S. Pat. No. 5,057,422).

Other suitable hosts for expressing full size or modified E1 endoglucanase include members of the genera: Trichoderma, Fusarium, Penicillium, Bacillus, Xanthomonas, Streptomyces, Aspergillus and Pichia, for example. Some of these microorganisms also serve as sources of endoglucanase genes for the formation of mixed domain genes for the production of hybrid enzymes.

Expression of the full size E1 endoglucanase gene has been demonstrated in E. coli, Pichia pastoris and in Streptomyces lividans.

Expressing full size or modified E1 endoglucanase in E. coli may be performed under control of a T7 bacteriophage promoter or other promotor recognizable by E. coli. Expression of full size E1 in E. coli has been enhanced considerably relative to the native gene with the constructs of the present invention. Expression of the full size E1 endoglucanase coding sequence in S. lividans has been achieved twice more with several different constructs employing two different promoters. These are the tipA promoter (thiostrepton-inducible) and the ST I-II promoter isolated from a trypsin inhibitor gene from Streptomyces longisporus. Expression levels of active, full size, secreted E1 endoglucanase up to 20 mg/L have been achieved with the ST I-II promoter.

Expression of the full size or modified E1 endoglucanase coding sequence in the filamentous fungi, Aspergillus niger, A. awamori, A. oryzae, A. terreus and/or A. nidulans and Trichoderma reesei is achievable using various promoters derived from Aspergillus or Trichoderma. These promoters include, but are not limited to, G3PDH (glyceraldehyde-3-phosphate dehydrogenase), glucoamylase, β -tubulin and IPNS (isopenicillin N synthase) from Aspergillus and CBH I (cellobiohydrolase I), alcohol dehydrogenase, triosephosphate isomerase and α -amylase from T. reesei.

The Acidothermus cellulolyticus full size E1 endoglucanase gene was cloned, as described in Ser. No. 08/276,213, and expressed in Pichia pastoris using the AOX1 promoter and 3' sequences. This expression is taught in Example 1.

EXAMPLE 1

Expression of the Entire Full Size E1 Gene in *Pichia pastoris*

Expression of active E1 endoglucanase in the range of 0.75-1.5 g/L has been accomplished in the yeast, Pichia

4

pastoris, by splicing the methanol-inducible alcohol oxidase (AOX1) promoter, including the signal sequence from the *P. pastoris* alcohol oxidase polypeptide to the mature coding sequence of the E1 endoglucanase gene.

P. pastoris has been shown to be a useful host organism 5 for expression of large quantifies of diverse heterologous proteins. P. pastoris was used to express large quantities of active full size E1.

Plasmid 4-5, a pGEM-7 (Promega Corp.) derivative carrying a 3.7 kb genomic fragment of Acidothermus cellulolyticus DNA and harboring the entire E1 gene, was used as a template in PCR reactions to amplify the E1 coding sequence for subsequent cloning into the Pichia secretion vector, pPIC9. The primer annealing to the non-coding strand of the template, "E1-f", is a 30-mer with 18 bases of homology to the template, beginning at the mature N-terminus of the E1 polypeptide. The segment of E1-f which is not homologous to the template molecules incorporates 4 codons which encode the C-terminal 4 amino acids of the αF signal peptide which is present in pPIC9 and also includes an XhoI site at the 5' end for use in subsequent cloning into pPIC9. The primer annealing to the coding strand and priming synthesis of the non-coding strand, "E1-r", is a 24-mer with 18 bases of homology to the template. The sequence of E1-r corresponds to the last 5 codons of the E1 coding sequence and the stop codon. The non-homologous 5' tail of E1-r adds an AvrII restriction site for use in subsequent cloning into pPIC9.

6

E1 on Western blots. Some of the reactive material on these blots runs as two diffuse high molecular weight bands. This material may be heavily glycosylated relative to the main band on the blots, which runs at a molecular weight only slightly higher than native E1 (75–80 kDa vs. 72 kDa). Most of the E1 produced in these cultures was secreted into the medium, as intended. The activity of the E1 secreted into the medium was demonstrated by the ability of these crude culture filtrates to hydrolyze 4-methylumbelliferyl- β -D-cellobioside (MUC), whereas control culture supernatants did not hydrolyze MUC.

Fed-batch fermentations of one of the Pichia transformants were conducted over a period of 4 days. Cultures were grown to a high optical density on 4% glycerol (w/v). When glycerol was exhausted (approx. 30 hours), methanol was fed as the sole carbon source over a period of 66 additional hours. After purification of the E1 from a portion of this culture, the yield was estimated at 1.5 g/L.

The present invention also comprises a further improvement of the E1 enzyme by modification of the physical structure of the enzyme to enhance its thermostable properties. At 80° C., the modified enzyme of the present invention demonstrates increased stability relative to the parent, full size E1, as well as an approximately 10° C. increase in its optimal temperature for activity.

This enhancement is brought about by the cleavage of the cellulose binding domain from the catalytic domain of the full size E1 enzyme. The cleavage of the cellulose binding

Cloning of the E1 expression construct in pPIC9 was 40 accomplished using the Pichia Expression Kit supplied by Invitrogen (San Diego, Calif.). All procedures are those recommended by Invitrogen. The PCR reaction produces a 1584 bp fragment containing the entire open reading frame for the full size, mature E1 polypeptide. The PCR product 45 was cloned directly into the TA vector, PCRII (Invitrogen, San Diego, Calif.). A single clone containing the PCR product was digested with XhoI and AvrII and the resulting 1.6 kb fragment was cloned into the same sites of pPIC9 to produce pPIC9-E1, which is diagrammatically represented 50 in FIG. 1. Several independent pPIC9-E1 isolates were screened for the existence of XhoI and AvrII sites and subsequently subjected to DNA sequencing across the signal peptide/E1 fusion junction to verify the correct sequence context in this region.

The pPIC9-E1 plasmid was linearized at the unique Stul site and transformed into spheroplasts of *P. pastoris* strain GS115. His⁺ transformants were selected for the Mut⁻ phenotype. Twenty independent His⁺Mut⁻ isolates were screened by PCR using the AOX1 primers. Clones which 60 displayed a 1.8 kb PCR product were screened for expression of E1 endoglucanase activity after growth on methanol.

The media and intracellular contents of cells from cultures grown in the presence of methanol for two days were screened by Western blot analysis using a monoclonal 65 antibody specific for the E1 endoglucanase. All P. pastoris clones containing the foreign DNA were shown to express

domain from the catalytic domain can be accomplished by more than one method. The first method of cleaving the cellulose binding domain from the catalytic domain is the enzymatic cleavage of the cellulose binding domain from the full-sized E1. This is taught in Example 2.

EXAMPLE 2

Method for the Production, Purification and Papain Cleavage of Full Size E1 Endoglucanase to Produce E1 CAT.

Full size E1 enzyme was recombinantly produced using S. lividans strain TK24 expressing E1-pIJ702 grown in 30 g/L Tryptic Soy Broth (Difco) with 5 ug/mL thiostrepton using a New Brunswick Microferm fermenter. The fermentation broth (10 L) was harvested using a CEPA continuous flow centrifuge, the supernatant concentrated and diafiltered against 20 mM Bis-Tris, pH 5.8 to a final volume of 300 mL using an Amicon CH2 concentrator and 10,000 MW cutoff hollow-fiber cartridges. Yields of native 72 kDa molecular weight E1 from these fermentations ranged from 1.2 to 2 mg/L, as estimated from purification yields.

The recombinant full size enzyme was purified with essentially a three step purification process consisting of hydrophobic interaction chromatography (HIC) followed by anion exchange and finally, size exclusion chromatography. The HIC step employed a column packed with 250 mL of Pharmacia Fast Flow Phenyl Sepharose. This was followed

by anion-exchange chromatography using a 6 mL Pharmacia Resource Q anion exchange column.

For the hydrophobic interaction step, ammonium sulfate was added to the concentrated culture supernatant of one 10 L fermentation to a final concentration 0.5M. A total volume 5 of 300 mL of this concentrate was loaded onto the phenyl sepharose column and washed extensively with 20 mM Tris. 0.5M (NH₄)₂SO₄ pH 8.0. The column was developed with a linear decreasing gradient (0.5M-zero) of (NH₄)₂SO₄. Recombinant full size E1 eluted at zero salt concentration. Fractions containing E1 activity were identified using 4-methylumbelliferyl β-D-cellobioside (MUG) assays. Active fractions were combined, concentrated and diafiltered against 20 mM Tris pH 8.0, after which they were loaded directly onto the anion-exchange column and eluted 15 with a increasing NaCl gradient (0-300 mM). A final buffer exchange and purification step was done using size exclusion chromatography with a 2.6 cm×10 cm Pharmacia Superdex 200 column and using one of two buffer systems; either 20 mM acetate, 100 mM NaCl, pH 5.0 buffer or 50 mM ammonium acetate, pH 6.2 buffer depending upon the eventual use of the enzyme.

The E1 catalytic domain was produced by proteolytic cleavage with papain. The proteolytic digestions were done in 50 mM ammonium acetate, pH 6.2 buffer. Molar ratios of 25 72 kDa MW E1/23 kDa MW papain cleaved E1 of 6/1 were used. Papain digestions were incubated at 28° C. for 24 h. The catalytic domain was separated from full size recombinant E1, E1-CBD (cellulose binding domain), linker peptides and papain by SEC using a 2.6 cm×10 cm Pharmacia 30 Superdex 200 column.

A chromatogram demonstrating the purification of modified catalytic domain from papain cleaved S. lividans recombinant E1 by size exclusion chromatography is shown in FIG. 4. Peak A is unmodified full size E1 enzyme, peak B is the catalytic domain, and peak C is the cellulose binding domain of the cleaved product. Also, peak B reacts with a monoclonal antibody (MAB) specific for full size E1, thereby confirming that this fragment contains epitopes for which this MAB shows specificity.

In order to confirm the identity of the peptide isolated from the papain digestion, the peak B peptide was subjected to analytical ultracentrifugation using a Beckman Optima XLa centrifuge. Sedimentation equilibrium analysis yielded an estimated molecular weight of 41,600 daltons for this peptide. SDS-PAGE analysis rendered a molecular weight estimate of 42,000 daltons. These values compare well with the calculated molecular weight predicted from amino acid sequence for the catalytic domain (i.e., 40,192 da.) Due to the limits of detection of the techniques utilized herein, these molecular weight values are well within the range of experimental error and therefore the molecular weight of the E1 CAT is in the range of about 40,000 to 42,000 Daltons.

Comparison of the thermal stability at 80° C. of pE1 CAT and full size E1 endoglucanase, both produced from a Streptomyces host, can be seen in FIG. 2. The enzyme was

activity using 1 mg/mL p-nitrophenyl-β-D-cellobioside substrate in 20 mM acetate, 100 mM NaCl, pH 5.0, at 65° C. for 30 minutes. Activity values are expressed as a percentage of the activity detected at time zero (e.g., 100%).

Comparison of the temperature optima of modified E1 and full size E1 was performed using 1 mg/mL p-nitrophenyl-g-D cellobioside substrate in 20 mM acetate, 100 mM NaCl, pH 5.0, at various temperatures with a 30 minute incubation time. See FIG. 3. The temperature optimum of the E1 CAT produced by papain cleavage is increased by 10° C. relative to the full size E1. Activity values in FIG. 3 are expressed as a percentage of the maximum activity detected.

It is further contemplated that one may include more than one catalytic domain in the hybrid enzyme. This may allow for a further increase in specific activity. Also, a catalytic domain containing cellulase activity other than endoglucanase activity may be included as well to reduce the number of cellulase enzymes one needs to add to a cellulosic substrate for polymer degradation.

EXAMPLE 3

Production of E1 CAT by Genetic Truncation of the Full Size E1 Coding Sequence

The amino acid sequence of E1 CAT is shown in FIG. 5. The entire amino acid sequence of pE1 CAT, including the C-terminus, is confirmed by the x-ray crystal structure derived from this molecule. An alternative to production of E1 CAT by papain cleavage of full size E1 is taught A molecular genetic approach was also developed to produce E1 CAT.

A strategy was desired to generate a genetically truncated E1 gene which would produce E1 CAT without requiring any downstream processing to achieve E1 CAT from a precursor molecule, as for papain cleavage. One way to accomplish this is to introduce a translational stop codon at or near the C-terminal residue of the catalytic domain.

The 2.3 kb Bam H1 fragment containing most of the E1 gene was subcloned into pAlter-1 in preparation for sitedirected mutagenesis (pYCC100). The nucleotide sequence for this fragment is shown in FIG. 6. A mutagenic 36-mer oligonucleotide (underlined in FIG. 6) was synthesized (5'-ATTTTCGATC CTGTCTAATG ATCTGCATCG CCTAGC-3')(SEQ ID NO:5). Using the Altered Sites® II kit (Promega Corp.) two consecutive codons immediately downstream of the C-terminal residue of E1-CAT (as determined by x-ray crystallography) were changed to different stop codons (TAA, TGA). The six mutagenic nucleotides are double-underlined in FIG. 6. The DNA sequence of this clone (pYCC101) has been confirmed by dideoxy DNA sequencing in the region of the site-directed mutations using the T7 Sequenase kit supplied by US Biochemical. (Cleveland, Ohio). A plasmid map of pYCC101 is shown in FIG. 7.

Native AA seq Native DNA seq Mutagenic oligo Mutated DNA seq Mutated AA seq S I F D P V G A S A S P S S Q (SEQ ID NO: 6)
TCGTCGATTTTCGATCCTGTCGGCGCGTCTGCATCGCCTAGCAGTCAA (SEQ ID NO: 7)
ATTTTCGATCCTGTCTAATGATCTGCATCGCCTAGC (SEQ ID NO: 8)

TCGTCGATTTTCGATCCTGTC TAATGATCTGCATCGCCTAGCAGTCAA (SEQ ID NO: 9)
S S I F D P V . S A S P S S Q (SEQ ID NO: 10)

incubated at 80° C. in 20 mM acetate, 100 mM NaCl, pH 5.0 buffer with timed aliquots removed, and each assayed for

Mutagenized DNA was transformed into E. coli strain ES1301. Transformants were screened for resistance to

65

ampicillin and sensitivity to tetracycline in order to identify clones carrying the putatively mutagenized E1 gene. Many ampicillin-resistant candidate clones were subsequently screened on plates containing 1 mM 4-methylumbelliferylβ-D-cellobioside (MUC) to verify expression of active E1. Plasmid DNA was prepared from several clones and employed as templates in dideoxy DNA sequencing reactions using the Sequenase® kit (U.S. Biochemical, Cleveland, Ohio) to verify the sequence of E1 DNA in the region of the intended mutation. The mutated sequence was 10 detected in every clone which was sequenced. One of these clones was selected and designated pYCC101. Each of the successfully mutated clones expresses a protein not present in control cells and which migrates at a molecular weight of approximately 42 kDa in SDS-PAGE gels. This 42 kDa 15 protein also reacts with a monoclonal antibody specific for the E1 endoglucanase on Western blots, thus confirming its identity as E1 CAT.

EXAMPLE 4

Differential Scanning Calorimetry (DSC)

Calorimetric studies of the denaturation of the full size E1 enzyme and the proteolytically cleaved E1 CAT were carried at pH 5.0 in 50 mM sodium acetate, using a Microcal MC-2 differential scanning microcalorimeter over a temperature range of 25°-95° C. and using a scan rate of 20° C./h. For the examples shown in FIG. 2, the protein concentrations were 0.24 mg/mL for the native E1 enzyme and 0.14 mg/mL for E1 CAT.

FIG. 8 shows the DSC thermogram for full size E1 enzyme displaying a prominent denaturation endotherm peak at approximately 78° C. This compares to the thermogram of E1 CAT, which shows no peak below at least 88° C. Since the thermograms are uninterpretable above 88° C. (limitation of the instrumentation), a reasonable conclusion is that the denaturation endotherm for the catalytic domain lies somewhere above 88° C.

EXAMPLE 5

Method for the Determination of the Functional Half-life of E1 and E1 CAT

The time at which haft of the original endoglucanase 45 activity remains following pre-incubation at 80° C. is

referred to as its functional half-life. Enzymes were preincubated at 80° C. in 20 mM acetate, 100 mM NaCl, pH 5.0, in concentrations of 13.14 µg/mL for the full size enzyme E1 and 19.53 µg/mL for E1 CAT. Small aliquots (100 µL) from single tubes were removed at various limes and assayed for activity by adding the 100 µL enzyme aliquot to 700 µL 20 mM acetate, 100 mM NaCl, pH 5.0, and 200 µL of 5 mg/mL p-nitrophenyl \(\beta\)-D cellobioside. The assay mixture was incubated at 65° C. for 30 minutes and the reaction stopped by adjusting the pH by addition of 2 mL of 1M Na₂CO₃. The activity was then measured by determining the concentration of the nitrophenolate anion, released as a result of catalytic activity of the enzyme, by measuring the absorbance at 410 nm of the quenched samples. Results shown in FIG. 2 demonstrate that E1 CAT has a functional half-life at 80° C. which nearly doubles that of full size E1 (16 h vs. 9 h).

Other options for generating genetically truncated E1 contemplated by the present inventors to be part of this invention include PCR of the coding sequence incorporating a non-homologous stop codon into the downstream synthetic primer; or using available restriction sites downstream of the DNA encoding the catalytic domain to delete the DNA sequences encoding the linker peptide and cellulose binding domain. Unless specifically defined otherwise, all technical or scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials have been described.

The foregoing description of the specific embodiments reveal the general nature of the invention so that others can, by applying current knowledge, readily modify and/or adapt for various applications such specific embodiments without departing from the generic concept, and, therefore, such adaptations and modifications should and are intended to be comprehended within the meaning and range of equivalents of the disclosed embodiments. It is to be understood that the phraseology or terminology employed herein is for the purpose of description and not of limitation.

All references mentioned in this application are incorporated by reference.

SEQUENCE LISTING

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

-continued

CTC GAG AAA AGA GCG GGC GGC TAT TGG 3 0 (2) INFORMATION FOR SEQ ID NO: 2: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 10 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: Protein (i x) FEATURE: (A) NAME/KEY: E1-f primer (x i) SEQUENCE DESCRIPTION: SEQ ID NO: 2: Leu Glu Lys Arg Ala Gly Gly Gly Tyr Trp (2) INFORMATION FOR SEQ ID NO: 3: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic Acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear (i i) MOLECULE TYPE: DNA (i x) FEATURE: (A) NAME/KEY: Elr (\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO: 3: CCT AGG TTA ACT TGC TGC GCA GGC 2 4 (2) INFORMATION FOR SEQ ID NO: 4: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 5 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: Protein (i x) FEATURE: (A) NAME/KEY; E1r (x i) SEQUENCE DESCRIPTION: SEQ ID NO: 4: Ser Ala Ala Cys Ala (2) INFORMATION FOR SEQ ID NO: 5: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic Acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear (i i) MOLECULE TYPE: DNA (\mathbf{x} \mathbf{i}) SEQUENCE DESCRIPTION: SEQ ID NO: 5: ATTTTCGATC CTGTCTAATG ATCTGCATCG CCTAGC 36 (2) INFORMATION FOR SEQ ID NO: 6: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 16 amino acids (B) TYPE: amino acid

-continued

	(C) STRANDEDNESS: single (D) TOPOLOGY: linear
(ii) N	MOLECULE TYPE: Protein
(x i) S	EQUENCE DESCRIPTION: SEQ ID NO: 6:
Ser Ser]	lle Phe Asp Pro Val Gly Ala Ser Ala Ser Pro Ser Ser Gla 5 10 15
(2) INFORMATI	ON FOR SEQ ID NO: 7:
(i)S	EQUENCE CHARACTERISTICS: (A) LENGTH: 48 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
(i i) N	AOLECULE TYPE: DNA
(x i) S	EQUENCE DESCRIPTION: SEQ ID NO: 7:
CGTCGATT	TT TCGATCCTGT CGGCGCGTCT GCATCGCCTA GCAGTCAA 48
(2)INFORMATION	ON FOR SEQ ID NO: 8:
(i)S	EQUENCE CHARACTERISTICS: (A) LENGTH: 36 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
(ii) M	OLECULE TYPE: DNA
(i x) E	EATURE: (A) NAME/KEY: mutagenic olige
(x i) S	EQUENCE DESCRIPTION: SEQ ID NO: 8:
ATTTTCGAT	CC CTGTCTAATG ATCTGCATCG CCTAGC 36
2) INFORMATION	ON FOR SEQ ID NO: 9:
(i) Si	BQUENCE CHARACTERISTICS: (A) LENOTH: 48 base pairs (B) TYPE: mucleic scid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
(i i) M	OLECULE TYPE: DNA
(i x) F	EATURE: (A) NAME/KEY: mutated DNA
(x i) S	EQUENCE DESCRIPTION: SEQ ID NO: 9:
CGTCGATT	TT TCGATCCTGT CTAATGATCT GCATCGCCTA GCAGTCAA 48
2)INFORMATIO	ON FOR SBQ ID NO: 10:
(i) SI	EQUENCE CHARACTERISTICS: (A) LENOTH: 16 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
(ii)M	OLECULE TYPE: Protein
(ix)F	EATURE: (A) NAME/KEY: Mutated amino acid
(x i) SI	EQUENCE DESCRIPTION: SEQ ID NO: 10:
er Ser I	le Phe Asp Pro Val Xaa Xaa Ser Ala Ser Pro Ser Ser Gla

-continued

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 358 amino acids
 - (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: Protein
- (i x) FEATURE:

(A) NAME/KEY: E1-CAT

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

A 1 a	G 1 y	Gly	Gly	T y r	Trp	His	Thr	Ser	G 1 y	Агд	G 1 u	110	Leu	A s p 1 5	Ala
Asn	Asn	Val	Pro 20	Val	Arg	I 1 e	A 1 a	Gly 25	I I e	Asn	Тгр	P h e	G 1 y 3 0	Phe	Glu
Thr	Сув	A s n 3 5	Туг	Val	Val	H i s	G 1 y 4 0	Leu	Trp	Ser	Arg	A s p 4 5	Туг	Arg	Ser
Met	Leu 50	Asp	Gln	116	Lys	S e r 5 5	Leu	Gly	Туг	Asn	Thr 60	11 e	Агд	Leu	Рго
T y r 6 5	Ser	A s p	Asp	lle	L e u 7 0	Lys	Рго	Gly	Thr	Me t 75	Pro	Asn	Ser	1 I e	A s n 8 0
Phe	Туг	Gln	Met	A s n 8 5	Gln	Asp	Leu	Gln	G 1 y 9 0	Leu	Thr	Ser	Leu	G 1 n 9 5	V a 1
Met	Аsр	Lys	I 1 e 1 0 0	Val	Ala	Туг	Ala	G 1 y 1 0 5	Gln	I 1 e	G 1 y	Leu	Arg 110	I i e	Ile
Leu	Азр	Arg 115	His	Агд	Pro	Asp	C y s 1 2 0	Ser	G 1 y	Gin	Ser	A 1 a 1 2 5	Leu	Trp	Туг
Thr	S e r 1 3 0	Ser	V a 1	Ser	Giu	A 1 a 1 3 5	Thr	Тгр	Ile	Ser	A s p 1 4 0	Leu	Gln	Ala	Leu
A 1 a 1 4 5	Gla	Arg	Туг	Lув	G 1 y 1 5 0	Asn	Pro	Thr	Val	V a I 1 5 5	G 1 y	Phe	Asp	Leu	His 160
Asn	Glu	Pro	His	A s p 1 6 5	Pro	Ala	Сув	Trp	G1 y 170	Сув	Gly	Asp	Рго	Ser 175	Ile
Asp	Тгр	Arg	Leu 180	Ala	Ala	Glu	Arg	Ala 185	G 1 y	Asn	Ala	V a 1	Leu 190	Ser	Val
Asn	Рго	Asn 195	Leu	L e u	lle	Phe	V a 1 2 0 0	Glu	G 1 y	V a 1	G 1 n	Ser 205	Туг	Авп	G l y
Авр	S e r 2 1 0	Туг	Trp	Trp	GIy	G I y 2 1 5	Азп	Leu	Gln	Gly	A 1 a 2 2 0	Gly	Gln	Туг	Pro
V a 1 2 2 5	Val	Leu	Asa	V a 1	Pro 230	Asn	Arg	Leu	Val	T y r 2 3 5	Ser	Ala	His	Авр	Tyr 240
Ala	Thr	Ser	V a 1	T y r 2 4 5	Рго	Gla	Thr	Тгр	Phe 250	Ser	Азр	Рго	Thr	Phe 255	Pro
Asn	Авп	Met	Pro 260	Gly	I 1 e	Trp	Asn	L y s 2 6 5	Asn	Тгр	G 1 y	Туг	Leu 270	Phe	Asn
G 1 n	Asn	I 1 e 2 7 5	Ala	Pro	V a 1	Trp	Leu 280	Gly	Glu	Phe	Gly	Thr 285	Thr	Leu	G 1 n
Ser	Thr 290	Thr	Авр	Gln	Thr	Trp 295	Leu	Lys	Thr	Leu	V a 1 3 0 0	GIn	Туг	Leu	Arg
Pro 305	Thr	Ala	G 1 n	Туг	G 1 y 3 1 0	Ala	Asp	Ser	Phe	G 1 n 3 1 5	Тгр	Thr	Phe	Тгр	Ser 320
Тгр	Asn	Pro	Asp	S c r 3 2 5	G 1 y	Asp	Thr	GIy	G 1 y 3 3 0	I 1 e	Leu	Lys	Asp	A s p 3 3 5	Тгр
G 1 n	Thr	V a 1	A s p	Thr	V a l	Lys	Asp	G 1 y	Туr	Leu	A 1 a	Pro	I 1 e	Lys	Ser

-continued

3 4 5

340
Ser Ile Phe Asp Pro Val

(2) INFORMATION FOR SEO ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2293 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: DNA
- (i x) FEATURE:
 - (A) NAME/KEY: E1-CAT
- (x i) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGATCCACGT TGTACAAGGT CACCTGTCCG TCGTTCTGGT AGAGCGGCGG GATGGTCACC 6.0 CGCACGATCT CTCCTTTGTT GATGTCGACG GTCACGTGGT TACGGTTTGC CTCGGCCGCG 120 ATTTTCGCGC TCGGGCTTGC TCCGGCTGTC GGGTTCGGTT TGGCGTGGTG TGCGGAGCAC 180 GCCGAGGCGA TCCCAATGAG GGCAAGGGCA AGAGCGGAGC CGATGGCACG TCGGGTGGCC 2 4 0 GATGGGGTAC GCCGATGGGG CGTGGCGTCC CCGCCGCGGA CAGAACCGGA TGCGGAATAG 300 GTCACGGTGC GACATGTTGC CGTACCGCGG ACCCGGATGA CAAGGGTGGG TGCGCGGGTC 360 GCCTGTGAGC TGCCGGCTGG CGTCTGGATC ATGGGAACGA TCCCACCATT CCCCGCAATC 4 2 0 GACGCGATCG GGAGCAGGGC GGCGCGAGCC GGACCGTGTG GTCGAGCCGG ACGATTCGCC 480 CATACGGTGC TGCAATGCCC AGCGCCATGT TGTCAATCCG CCAAATGCAG CAATGCACAC 5 4 0 ATGGACAGGG ATTGTGACTC TGAGTAATGA TTGGATTGCC TTCTTGCCGC CTACGCGTTA 600 CGCAGAGTAG GCGACTGTAT GCGGTAGGTT GGCGCTCCAG CCGTGGGCTG GACATGCCTG 660 CTGCGAACTC TTGACACGTC TGGTTGAACG CGCAATACTC CCAACACCGA TGGGATCGTT 720 CCCATAAGTT TCCGTCTCAC AACAGAATCG GTGCGCCCTC ATGATCAACG TGAAAGGAGT 780 ACGGGGGAGA ACAGACGGGG GAGAAACCAA CGGGGGATTG GCGGTGCCGC GCGCATTGCG 8 4 0 GCGAGTGCCT GGCTCGCGGG TGATGCTGCG GGTCGGCGTC GTCGTCGCGG TGCTGGCATT 900 GGTTGCCGCA CTCGCCAACC TAGCCGTGCC GCGCCGGCT CGCGCCGCGG GCGGCGGCTA 960 TTGGCACACG AGCGGCCGGG AGATCCTGGA CGCGAACAAC GTGCCGGTAC GGATCGCCGG 1020 CATCAACTGG TTTGGGTTCG AAACCTGCAA TTACGTCGTG CACGGTCTCT GGTCACGCGA 1080 CTACCGCAGC ATGCTCGACC AGATAAAGTC GCTCGGCTAC AACACAATCC GGCTGCCGTA 1140 CTCTGACGAC ATTCTCAAGC CGGGCACCAT GCCGAACAGC ATCAATTTTT ACCAGATGAA 1200 TCAGGACCTG CAGGGTCTGA CGTCCTTGCA GGTCATGGAC AAAATCGTCG CGTACGCCGG TCAGATCGGC CTGCGCATCA TTCTTGACCG CCACCGACCG GATTGCAGCG GGCAGTCGGC 1320 GCTGTGGTAC ACGAGCAGCG TCTCGGAGGC TACGTGGATT TCCGACCTGC AAGCGCTGGC 1380 GCAGCGCTAC AAGGGAAACC CGACGGTCGT CGGCTTTGAC TTGCACAACG AGCCGCATGA 1440 CCCGGCCTGC TGGGGCTGCG GCGATCCGAG CATCGACTGG CGATTGGCCG CCGAGCGGGC 1500 CGGAAACGCC GTGCTCTCGG TGAATCCGAA CCTGCTCATT TTCGTCGAAG GTGTGCAGAG 1560 CTACAACGGA GACTCCTACT GGTGGGGCGG CAACCTGCAA GGAGCCGGCC AGTACCCGGT 1620 CGTGCTGAAC GTGCCGAACC GCCTGGTGTA CTCGGCGCAC GACTACGCGA CGAGCGTCTA 1680 CCCGCAGACG TGGTTCAGCG ATCCGACCTT CCCCAACAAC ATGCCCGGCA TCTGGAACAA 1740 GAACTGGGGA TACCTCTTCA ATCAGAACAT TGCACCGGTA TGGCTGGGCG AATTCGGTAC 1860

-continued

GACACTGCAA	TCCACGACCG	ACCAGACGTG	GCTGAAGACG	CTCGTCCAGT	ACCTACGGCC	1860
GACCGCGCAA	TACGGTGCGG	ACAGCTTCCA	GTGGACCTTC	TGGTCCTGGA	ACCCCGATTC	1920
CGGCGACACA	GGAGGAATTC	TCAAGGATGA	CTGGCAGACG	GTCGACACAG	TAAAGACGG	1980
CTATCTCGCG	CCGATCAAGT	CGTCGATTTT	CGATCCTGTC	TAATGATCTG	CATCGCCTAG	2040
CAGTCAACCG	TCCCCGTCGG	TGTCGCCGTC	TCCGTCGCCG	AGCCCGTCGG	CGAGTCGGAC	2100
GCCGACGCCT	ACTCCGACGC	CGACAGCCAG	CCCGACGCCA	ACGCTGACCC	CTACTGCTAC	2 1 6 0
GCCCACGCCC	ACGGCAAGCC	CGACGCCGTC	ACCGACGGCA	GCCTCCGGAG	CCCGCTGCAC	2220
CGCGAGTTAC	CAGGTCAACA	GCGATTGGG	CAATGGCTTC	ACGGTAACGG	TGGCCGTGAC	2280
AAATTCCGGA	тсс					2 2 9 3

We claim:

- 1. A DNA having the nucleotide sequence of SEQ ID NO.
- 2. The DNA according to claim 1, encoding the amino acid sequence of SEQ ID NO. 11.
- 3. A vector carrying the DNA according to claim 1 and a vector sequence encoding either an origin of replication or an integration site for a host genome.
- 4. The vector according to claim 3 further comprising DNA encoding a signal sequence operably linked thereto.
- 5. The vector according to claim 3 further comprising exogenous regulatory sequences capable of causing expression of said DNA in a suitable host; wherein a microorganism comprising *Pichia pastoris* contains the vector.
- A DNA according to claim 1 encoding the catalytic domain of E1 endoglucanase.
- 7. A DNA according to claim 1 encoding the catalytic domain and the linker peptide of E1 endoglucanase.

- 8. The DNA according to claim 6 further comprising at least one domain from a cellulase gene other than E1 endoglucanase.
 - 9. The DNA according to claim 7 wherein the DNA encodes a protein having an endoglucanase activity.
 - 10. A microorganism containing the vector of claim 5.
- 11. A method for producing an endoglucanase having the amino acid sequence of claim 2, comprising enzymatic cleavage of the cellulose binding domain of full size El endoglucanase.
- 12. A method for producing an endoglucanase having the amino acid sequence of claim 2, comprising inserting at least one stop codon after the sequence encoding the catalytic domain of full size E1 endoglucanase so that the linker peptide and cellulose binding domain of E1 endoglucanase are not expressed in the expression product.
- 13. The DNA according to claim 2 wherein the DNA encodes a protein having an endoglucanase activity.

* * * * *